Geospatial Multistate Archive and Preservation Partnership (GeoMAPP)

APPX-based Archives Enterprise Manager (AXAEM)

December 20, 2011

version: Final

Contents

APPX-based Archives Enterprise Manager (AXAEM)

During the first grant period of GeoMAPP (2008-2009), each state became familiar with systems used by the other states. In Utah's case, their system was built in-house using a commercial application development tool called APPX. When North Carolina and Kentucky first saw this system, reaction was very positive. Through the course of the grant project, the idea grew that this system could be of benefit to other archives. To that end, Utah made their code open-source, gave the application a name (AXAEM, which stands for APPX-based Archives Enterprise Manager) and arranged to have it distributed through their vendor, APPX Software[1]. As part of the AXAEM project, the vendor also agreed to update the source code to make the functionality more generic and to remove or adapt parts that were Utah-specific.

During the second grant period (2010-2011), one of the tools specifically being evaluated for use with geospatial records has been AXAEM. Functionality has been developed in tandem with business requirements identified in GeoMAPP.

Key Features

AXAEM manages the work product of an archives which may include both records management and archival responsibilities. The APPX software platform adds to AXAEM functionality by providing institutions a method to quickly change and adapt the application for their needs. Both sets of features will be described.

AXAEM Capabilities

AXAEM has been developed by Utah over the course of a couple of decades and is comprised of several sub-applications, the main one of which is the bibliographic management system, which is based on *Describing Archives: A Content Standard*

[1] See www.appx.com and www.axaem.com. The contract between the vendor and the State of Utah allows them to market and sell services surrounding AXAEM, such as technical support and maintenance, training, and consulting/programming. All of the Utah-produced code is available for free under the open source license while the software platform remains a proprietary product and has some cost associated with it. At the time of this report, the cost for a permanent license was $300 per concurrent user and $1,200 for design rights (the ability to change programming). A variety of maintenance packages are available, both for the underlying software as well as AXAEM itself. Utah currently has a 20-user license, and 1 design right. Annual maintenance runs about $3,000 per year. Utah only pays maintenance on the underlying software, not AXAEM, since that is supported in-house. Consulting/programming from the vendor costs Utah $135/hour, which is based upon the maintenance contract. The amount of programming developed during the two GeoMAPP grant periods totaled $136,350, which does not include time spent programming by local staff.

(DACS)[2]. A single entry in the bibliographic system may be output as a retention schedule, a Machine Readable Cataloging (MARC[3])-encoded catalog record, or an Encoded Archival Description (EAD[4])-encoded finding aid. Multiple entries may be associated together as collections. A separate process within AXAEM associates record creators with all of their Encoded Archival Context—Corporate Bodies, Persons, and Families (EAC-CPF[5]) details. Supporting processes track microfilm, register records officers for training sessions, plus offer a web-based interface for public interaction to search for personal names or dates contained in record indexes and display digital collections of those records; harvest records in AXAEM via Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH[6]); view retention schedules, finding aids, microfilm reports, and agency history information through an agency name search; submit new retention schedules or edit existing ones; etc. Retention schedule data is shared with a third-party box inventory system, used for space management in the Utah state records center, and disposition dates are automatically calculated. Prior to GeoMAPP, AXAEM had no specific features to manage electronic records other than through retention schedules tied to media stored in the Archives. Those capabilities are now in place and described further below.

These processes are both well-established and constantly changing. Development tends to follow an organic path, depending on Utah's needs at the moment, as well as system requirements that spring from ever-changing professional standards and best practices.

This is how the AXAEM menu system appears to users currently:

[2] See http://www.archivists.org/governance/standards/dacs.asp.
[3] See http://www.loc.gov/marc/.
[4] See http://www.loc.gov/ead/.
[5] See http://eac.staatsbibliothek-berlin.de/.
[6] See http://www.openarchives.org/pmh/.

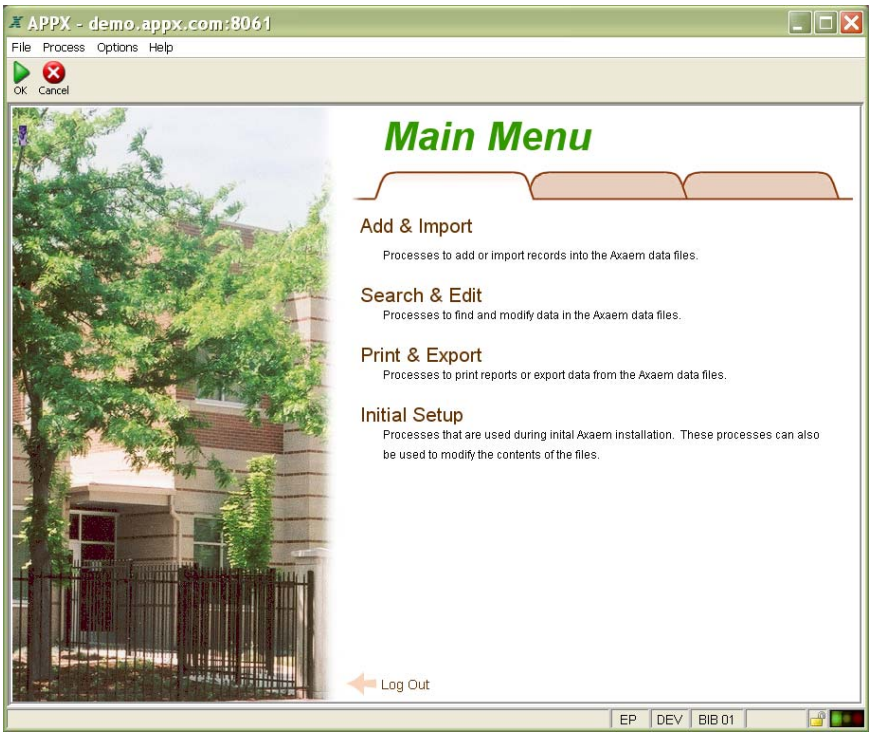


Figure 1: Main Menu

The main menu options are organized into four major categories of functionality. The following screenshots display the types of features available now as well as those still in the planning stage (or in the middle of being adapted for general use).

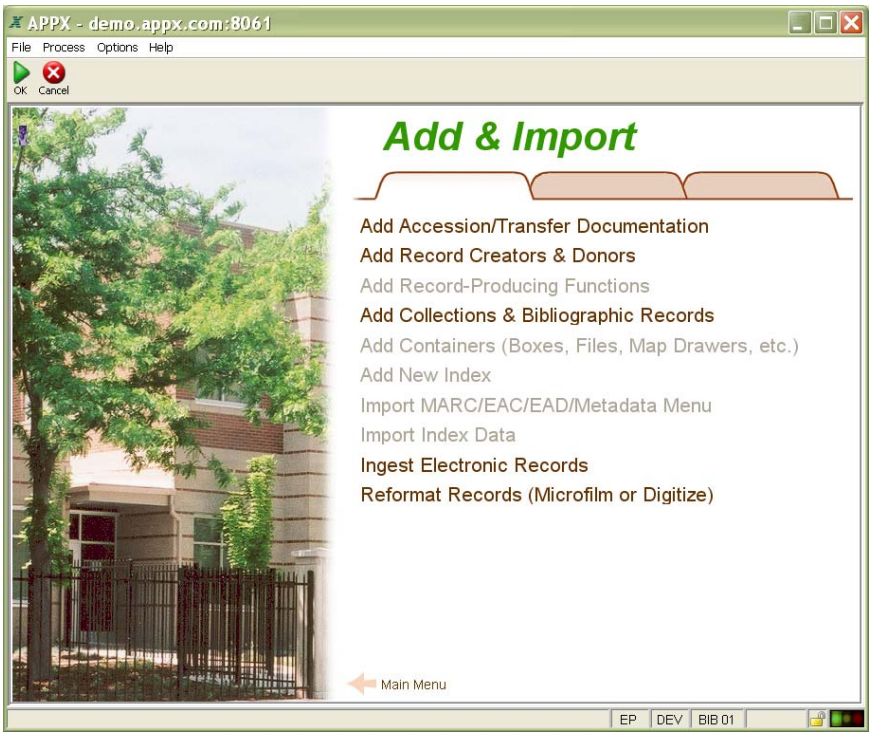


Figure 2: Add & Import menu

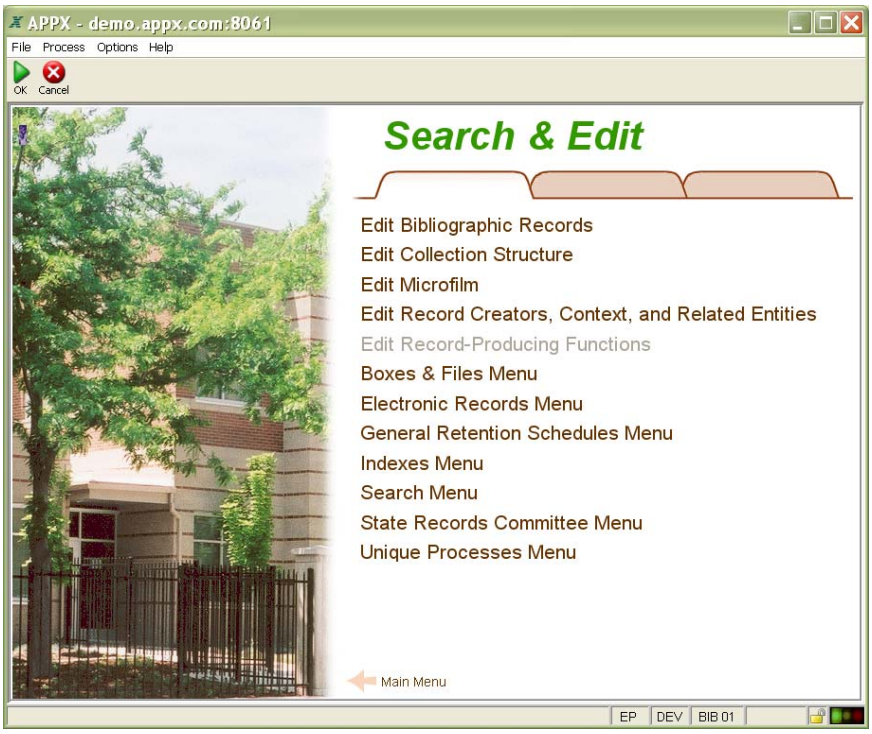


Figure 3: Search & Edit menu

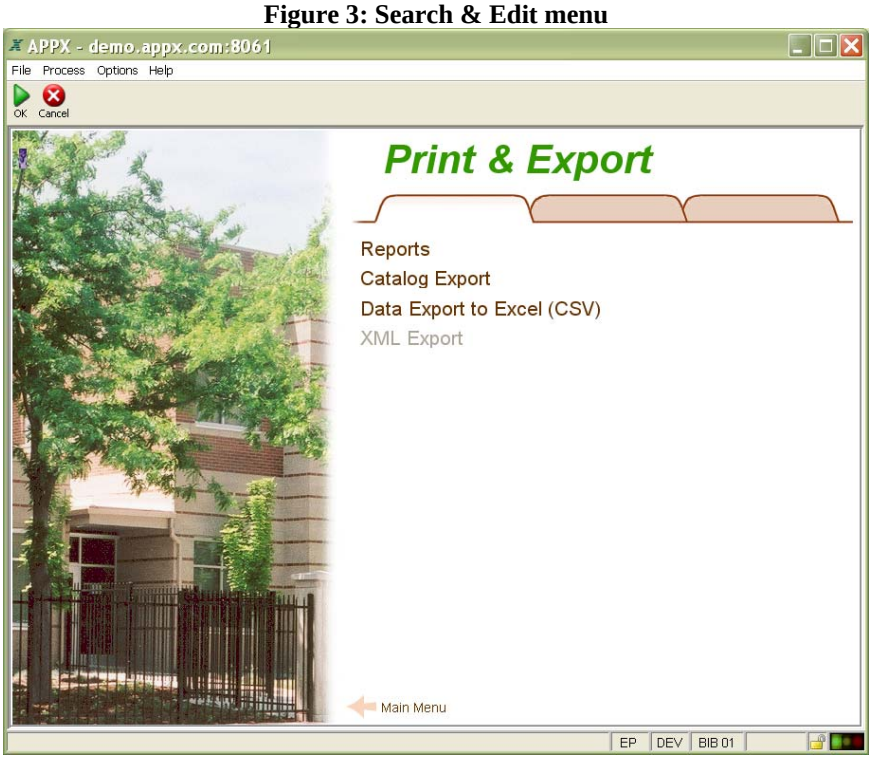


Figure 4: Print & Export menu

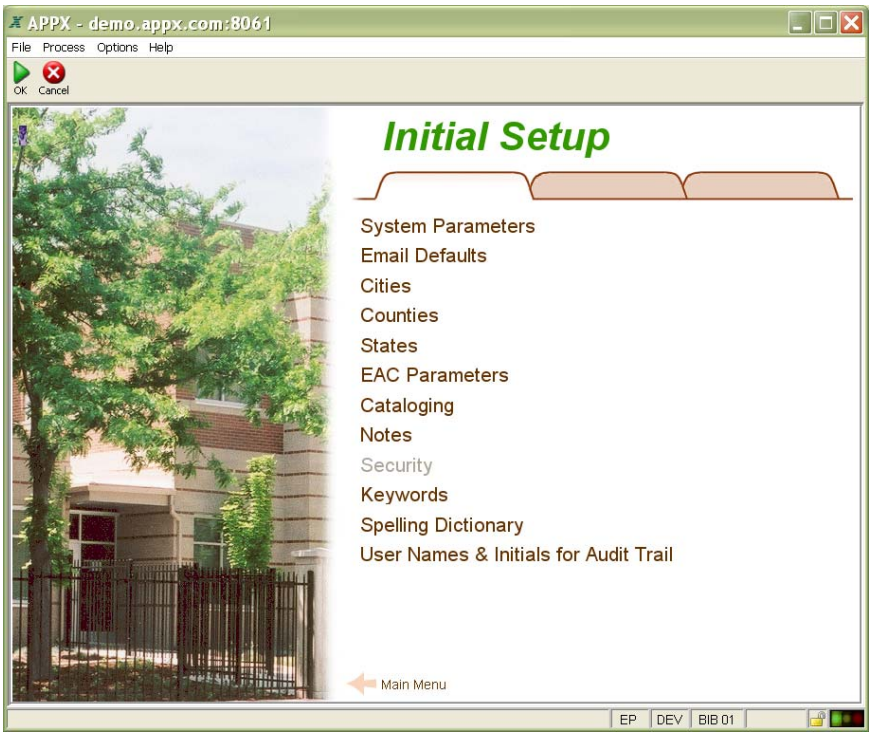


Figure 5: Initial Setup menu

From these menus, users can drill down to specific tasks, such as adding an entity or a series.

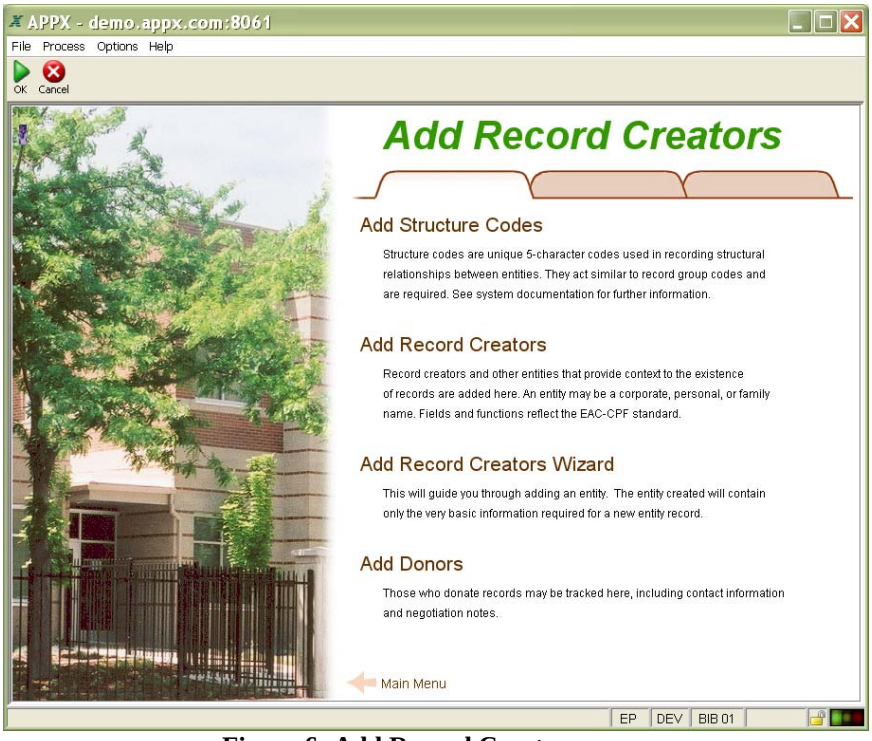


Figure 6: Add Record Creators menu

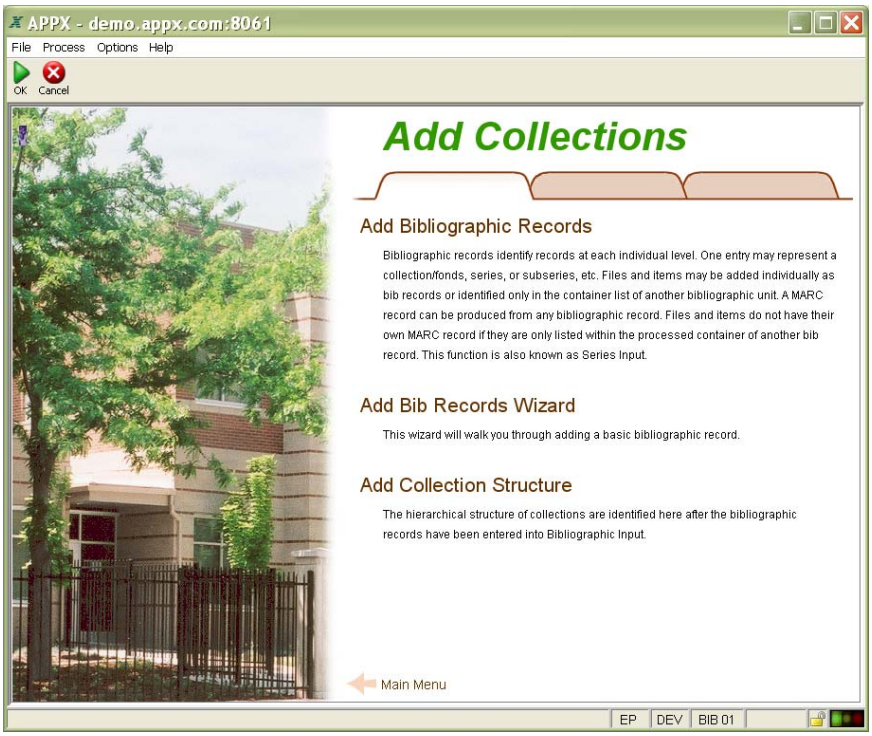


Figure 7: Add Collections menu

The menus lead the user to the end process they want to interact with, such as Bibliographic Input, or Entity Input.

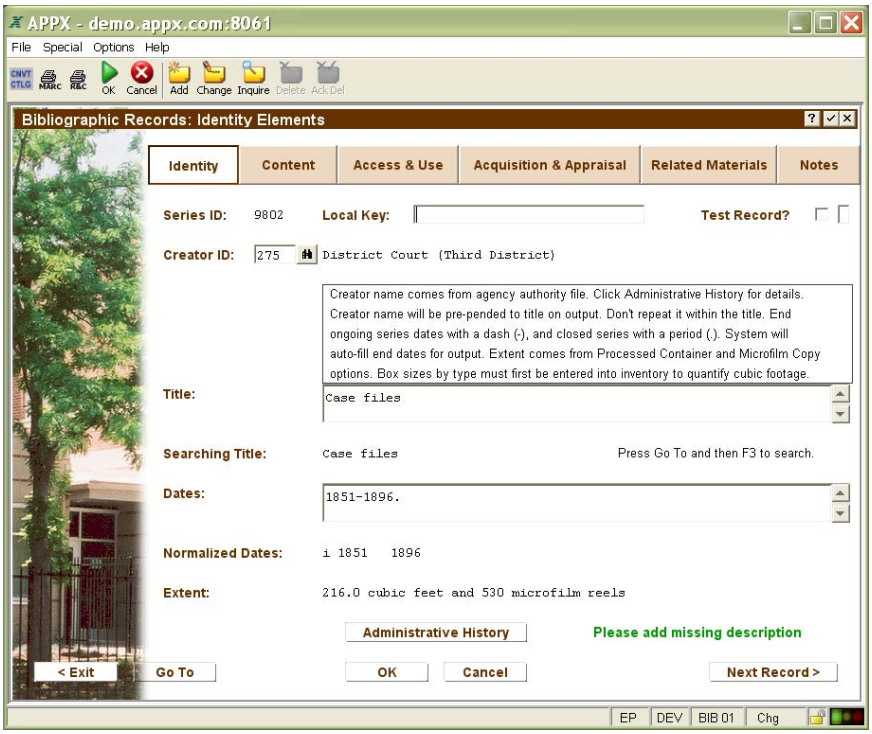


Figure 8: Bibliographic Input

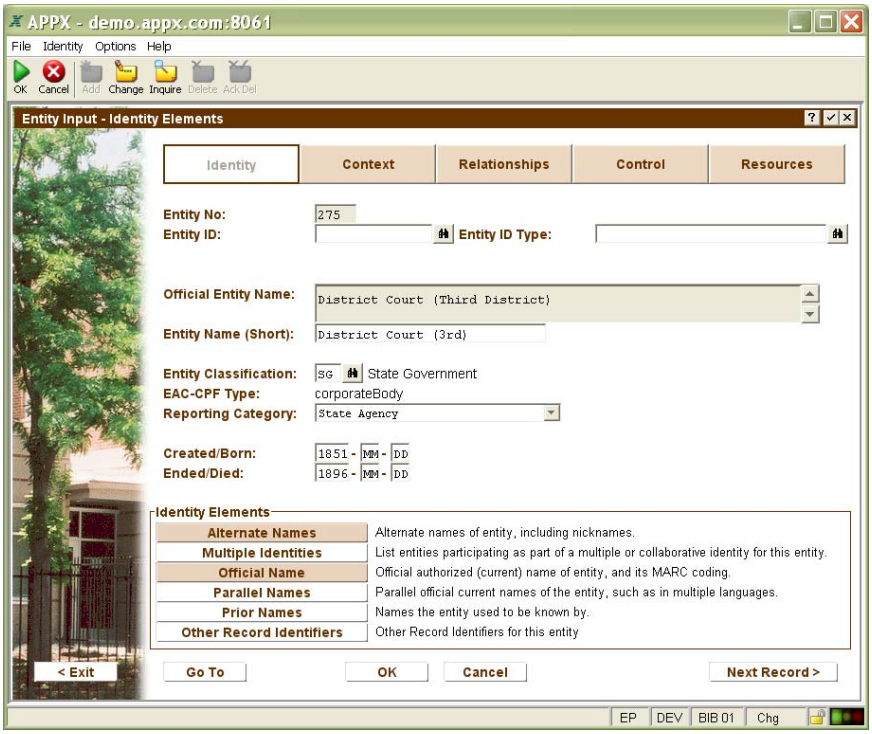


Figure 9: Entity Input

APPX Software Capabilities

APPX software uses a client-server TCP/IP architecture with a Java-based thin client. The APPX engine can be installed on Unix, Linux, or Windows operating systems, and the client can be installed and run locally on the APPX server as well as run remotely on Windows, Linux, and Macintosh desktops. The vendor also offers infrastructure-as-a-service hosting options via Amazon Web Services. Databases that are supported include APPX's native database called APPXIO, plus Oracle, and ODBC-connected PostGres, MySQL, and SQL Server. APPX can connect to more than one database simultaneously. In Utah's configuration, both APPXIO and Oracle are used. One instance of the application can talk to another instance of the application. It can integrate with other systems and launch server-side and client-side utilities, and since the base product is an application development tool, any feature in the application can be changed by the institution.

The nature of this tool automates traditional application development tasks. Utah's experience is that new database tables, fields, and input and output processes can be added within just a few minutes. More complicated functionality might take a few weeks to code by someone who is trained in the product. Writing the subroutine that automatically exports bibliographic data into an EAD finding aid took Utah three weeks to complete. This reduced the user's effort to export the EAD file to a single click.

End users that are designated System Administrators and/or Application Designers in their user rights can configure their login to start with the APPX design menu instead of the AXAEM menu, or access it by Canceling out of the AXAEM menu:

Figure 10: APPX Development Tool's main menu

Figure 11: APPX's Application Design menu

The Application Design menu in Figure 11 offers developers the ability to add or change tables and fields, as well as processes that users interact with, and ones that run behind the scenes. Code is added at event points (see Figure 12) defined inside these processes, such as when a screen is first displayed or after a record has been updated. These event points minimize the code needed to make an application function since processing instructions are handled by the APPX engine, leaving code developers managing only business rules. The APPX programming language is a fourth-generation language (4GL). Simple input and output processes can actually be created without necessitating any code at all. Just drag and drop fields to the screen where you want them and add graphical attributes to make it look pretty. Any process can call another process, allowing developers to reuse code such as subroutines and queries, minimizing the code required even further. In addition, processes external to APPX can be called and run, sharing data to and from APPX, which makes the external code behave as if it were natively part of the APPX application. This opens up the whole world of application libraries to potentially be used by APPX.

The data dictionary uses a concept called Domains, which is a field type that can be used as a pattern for similar fields. For example, if you have a bibliographic key (the unique ID used to identify a bibliographic record) referenced in several different tables, all of the attributes for the field (such as length) would be set within the domain, and if the developer needed to make a change to the field, only the domain settings would need to be changed. APPX will auto-update all the instances of the bibliographic key in all of the tables throughout the system.

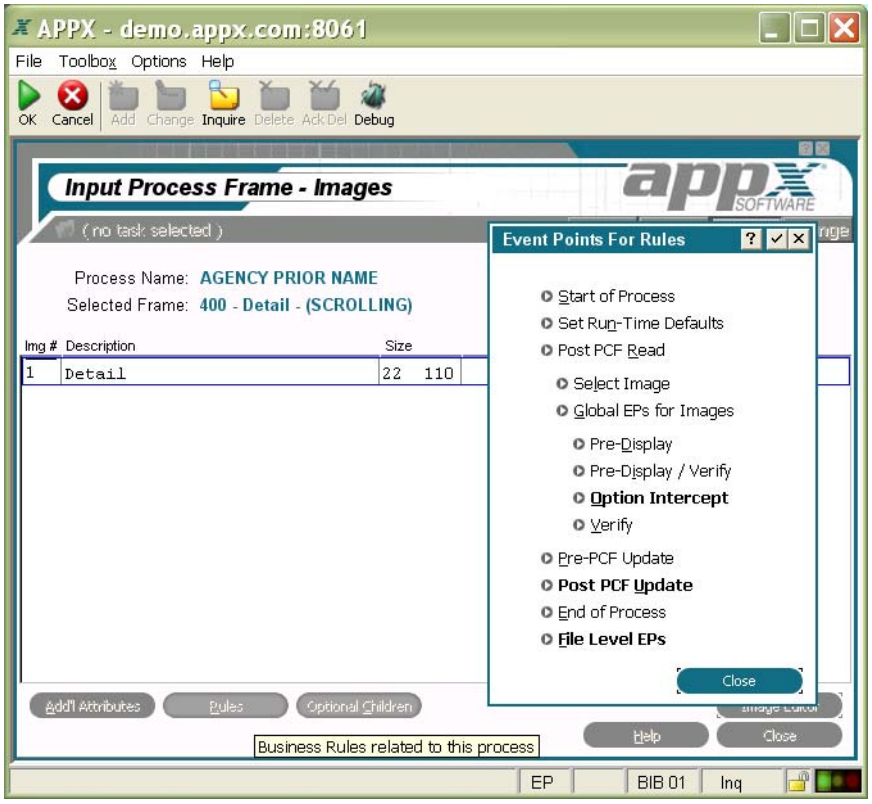


Figure 12: Event Point menu in an input process

Business rules are written using structured key word statements:

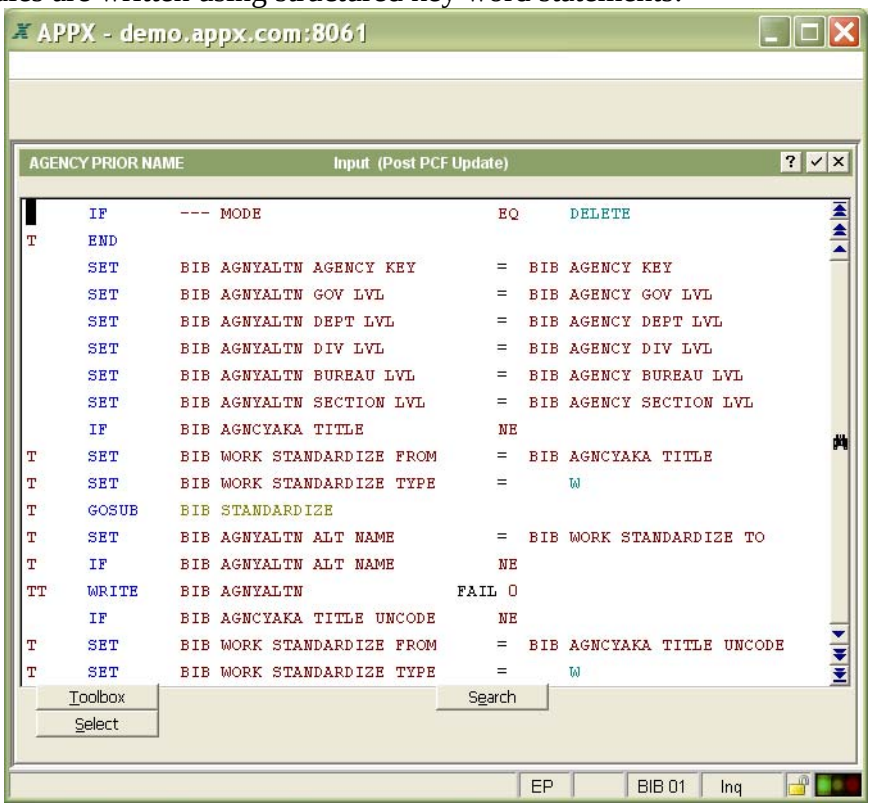


Figure 13: Code from an input process event point

The code in Figure 13 is taking the prior name of an agency, and writing it to the Agency Alternate Name file. It does this automatically every time a new prior name is added to the agency record. The Alternate Name file includes a copy of all versions of an agency's name, for use in searching.

One APPX installation can manage multiple versions of an application such as AXAEM. It can also manage separate databases tied to each version, so that an institution can have a test version of their data. This allows developers to have a design environment completely separate from the version being interacted with by users. When the developer is ready for system updates to take place, the new version can be copied over easily. A source code control option facilitates this by keeping track of everything that the developer changes and then, when ready, APPX can automatically transfer everything changed from the development applications to the production applications.

Alternatively, developers can make code changes in the same version that is being used in production. Since the APPX client provides both a design and run-time environment, these changes are automatically compiled as soon as someone accesses a process after the developer has exited the design view for that process. While changes to the database structure such as adding a field in a table that's already in use by someone requires users to be either logged out or in a menu for the restructure to take place, most application changes (including adding new tables) can be happening at the same time that users are interacting with the system. This allows developers to do their work right alongside users if desired. Utah has often discussed needed changes in a meeting, and an hour later those changes have already been made by staff members and the new features are up and running. While archivists could be trained to utilize some (or perhaps all) aspects of the design environment (Utah's application designer is an archivist, not a programmer), most likely applications would be developed by someone with formal programming training who has had the additional training required to understand the code structure used by APPX Software. Once trained, the effort required to produce results is significantly less than in traditional development environments.

Customized software may present a red flag for institutions that would rather use something out-of-the-box. AXAEM can be used either as a custom solution or highly-configurable off-the-shelf product. One thing that should be noted is that the APPX development tool has its own version and upgrade schedule than does an application built with the tool. From time to time the tool is upgraded with new features, and these features are accessible to applications written in APPX (such as AXAEM) without causing AXAEM itself to lose any of its look or functionality. However, AXAEM will have its own development path. The official version as released by the vendor will continue to be developed by the vendor and the open source community. If an institution takes the official version, installs it locally and then adapts their instance by adding new rules or changing labels to accommodate their needs, then those two versions of AXAEM will be out of sync, and upgrades of the official version that are applied to a local version could wipe out changes made locally unless steps are taken to avoid that through existing features, or new ones. At any rate, the issue related to retaining customizations through

AXAEM upgrades is one thing that the vendor will likely address formally. It is hoped that if an institution makes significant changes to their application, that those changes are made available to the official version so that all can benefit from the new development work.

Security is a prominent feature in APPX applications. Users can be tied to security groups, and rights are assigned to those groups. Security settings can control read, write, and execute permissions for processes, tables, and fields. Rights to add, delete, or change a record or field can be set individually, too, so if users are only allowed to add something, that is tightly controlled. In addition, the software offers data encryption. AXAEM has an additional security application that logs data changes and reasons for the change, and offers supervisors the opportunity to give permission for those changes to be made if the user doesn't already have rights to edit a field.

APPX allows developers to write an application using a natural language such as English, and use the Language Translation Tool to make the application appear in any other language. The translation table between the languages must be provided by the developer, but that means if the user's native language is German or Chinese, all of the labels, buttons, and system instructions will appear to the user in that language.

Report writing is another feature available. Generally, reports are created by application developers. Users can run and save pre-designed reports as either a text or PDF file. The PDF reports can include many graphical features, including charts derived from the data source. Delimited data such as .csv files can also be exported as a pre-designed output format. However, a report writer feature is available to users needing simple output or .csv files that exports data of their choosing. Printing can take place on either a networked or local printer.

APPX can interact with the Internet in a number of different ways. It can utilize .cgi data from a web form, integrate with a web search engine such as Solr, perform REST-ful web services such as OAI-PMH harvesting, or make the client launch when a user clicks a link on a website. The client, once launched, may request a login and password (or not), and go to a designated place such as the main menu or even a specific record on one specific screen.

Managing Electronic Records and Their Associated Metadata

GeoMAPP discussions led to realizations that new features must be added to AXAEM to account for the special needs of geospatial records in particular, but also electronic records in general. During the first phase of the grant project, the focus was on creating retention schedules for geospatial records. While records were being inventoried in the field to produce these schedules, geospatial experts informed Utah archivists of the importance of specific metadata that is needed to use and understand the records. In response, Utah created a child table under the bibliographic record to extend the data model and to support the geospatial metadata requirements. The intent was to follow suit for other types of electronic records (audio, video, text, databases, etc.). The geospatial table supported the following data elements: creation and publication dates, datum, projection, scale, resolution, GIS-supplied filename, descriptive title, file size, file location, and file type (shapefile, geodatabase, pdf, etc.). Data was entered manually to this table as geospatial record series were inventoried and posted to Utah's FTP server. Later, EAD finding aids were programmatically generated from this metadata. The metadata did not need to be entered twice to produce the finding aid.

At some point between the first and second phase of the grant project, this data design was deemed inadequate. Two reasons contributed to this conclusion. The first was that a geospatial dataset is usually comprised of multiple files which must be managed collectively and share a description (in the case of shapefiles and geodatabases), but each file also needs to be managed individually, such as recording and checking its own checksum value. The second was that within geospatial project files, multiple file types could be part of the same work process, including word processing, spreadsheet, and image files. All of these would need to be identified separately for preservation and metadata purposes, and yet grouped together for descriptive and access purposes. At the same time, other users in Utah wanted an easy way to produce metadata for born-digital and imaged records for use in uploading digital files into the CONTENTdm[7] system.

The result was to create a new application for electronic records. A two-tier data design was developed comprised of 1) Electronic Records, which represented the individual files being managed, and 2) Object Groups, which represented collections of related individual items that represent a single managed intellectual entity, as well as the folder structure of ingested records. The Object Groups can also maintain their structural relationship to other related Object Groups and Electronic Records, allowing the user to write descriptive information at the top parent folder or any level beneath. Supporting tables for media types, format types (with preservation rules) and transfer/accession documentation also were added. For individual Electronic Records, due to the number of possible format and metadata standards that could apply to any one electronic file, many of those format-specific details are being captured as child data to the Electronic Record. That optimizes the data design so that only the relevant fields are populated for each managed item.

[7] See http://www.contentdm.org/.

The Electronic Records menu was extended to support the ability to manage the individual Electronic Records (item-level), as well as the Object Group managed entities. Each of these options allows the user to input the metadata relevant to the individual item or managed object. Figure 14 shows how the menu currently looks:

Figure 14: Electronic Records Menu

A significant aspect of any electronic records management system is the establishment of the data model representing the managed metadata. The initial metadata standard used as the basis for the fields in the electronic record was Dublin Core[8]. With the focus on geospatial data, support for Federal Geographic Data Committee (FGDC[9]) metadata was added.

After defining the metadata model, strategies and techniques must be identified to populate the metadata fields associated with the electronic records being managed. Local users expressed a strong desire to automate as much of the metadata entry as possible. In the past, this was done by integrating spreadsheets into the import/export process. Spreadsheets offer many auto-fill features, and users were comfortable with that methodology. At the same time, Utah was aware of tools that automate metadata capture by scanning files in a folder structure and extracting properties embedded within the files to an XML format. Two tools Utah explored were the metadata extractor from the National Library of New Zealand[10], and JHOVE2[11]. Later, Utah became aware of the

[8] See http://dublincore.org/ as well as the Mountain West Digital Library's metadata standard at http://mwdl.org/docs/MWDL_DC_Profile_Version_2.0.pdf which formed the basis of AXAEM field names and behaviors.
[9] See http://www.fgdc.gov/.
[10] See http://meta-extractor.sourceforge.net/.

tool FITS[12], which wraps the metadata extracted from several tools (including the NZ extractor, Droid, the original JHOVE, and various file utilities) into a single XML file. Since XML is increasingly a standard adopted by archives for every conceivable purpose, plus having descriptive metadata coming natively with records in the case of FGDC, the ability to parse XML data directly in AXAEM became a significant functional requirement. In response, APPX added parsing capabilities in April 2010. The feature was designed so that data from any XML file could be mapped to any field in AXAEM. These map definitions were then tied to advanced update processes that could further process or manipulate the imported data before it was copied into the metadata fields for the associated electronic file and object groups. The user interface required just one single click to complete the metadata extraction, metadata processing, and data ingest.

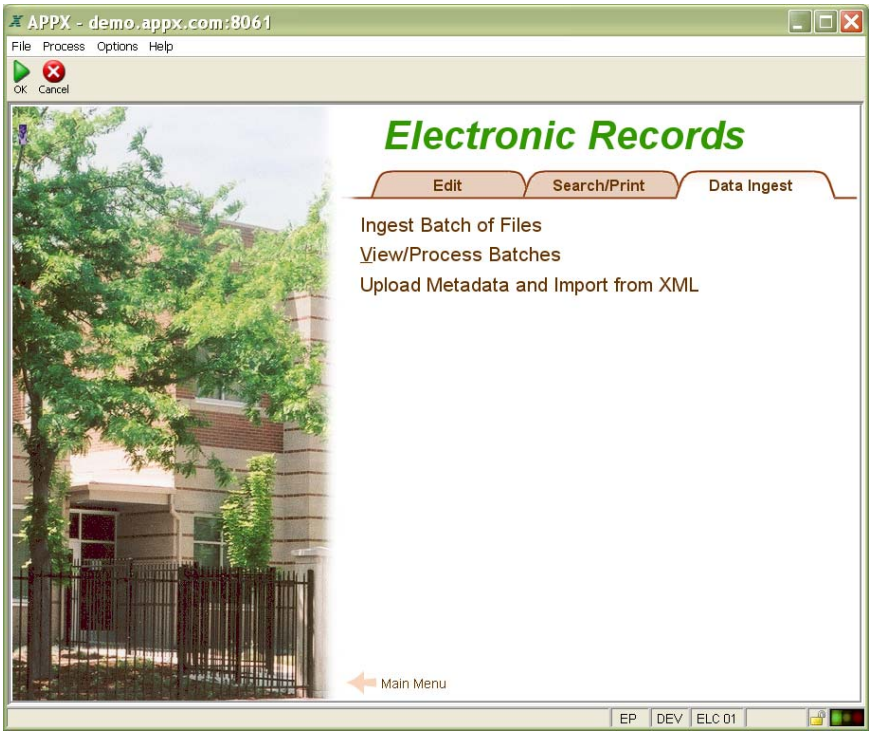


Figure 15: Electronic Records Menu, Data Ingest tab

Newest Features

The advancement of the data ingest process responded to the desire to ingest a single XML file filled with technical metadata, and a separate file with descriptive metadata, all as a single process. The AXAEM interface was then changed to allow the user to designate a folder, and the system would identify the file types and object groups it contained, incorporate geospatial metadata from a provided FGDC-compliant metadata file, calculate a checksum on each file, and write the record(s). In addition, the physical electronic record and its metadata needed to be linked together, and a process was added

[11] See https://bitbucket.org/jhove2/main/wiki/Home.
[12] See http://code.google.com/p/fits/.

to allow data ingest of the record by sending it to the server where the application has direct access to the file.

As of November 2011, the ingest process does the following:

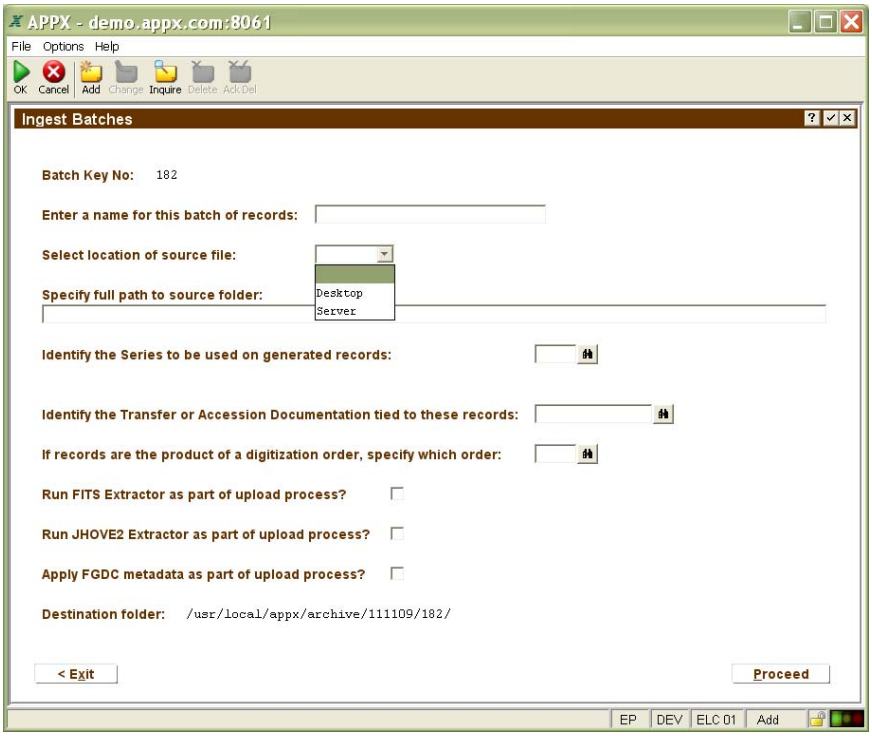


Figure 16: Data Ingest Screen

1. Users are prompted to provide the following information:
    a. Include a unique batch ingest name so that each Submission Information Package could be identified, reviewed, and acted upon later if needed.
    b. Point to a folder on either the desktop/networked drive or server where records to be ingested are currently stored.
    c. Identify the bibliographic ID that these records should be tied to.
    d. Identify the records transfer (accession) documentation the records should be tied to.
    e. Identify whether the records were produced from a digitization project, and if so, which one. This would tie the metadata from that process to the ingested records.
    f. Identify which metadata extractors the user wants to run. Currently, those choices include FITS and JHOVE2. Descriptive metadata that comes with the records could be presented in FGDC form or Dublin Core as exported from CONTENTdm. The content of all metadata sources is then written to the database.
2. The ingest then proceeds and does the following:
    a. A checksum is captured for each file before they are moved to the AXAEM server.

b. Files are transferred to the AXAEM server (or, if already on the server, copied to a designated storage location).
c. The checksum is run again and compared to its previous value to make sure that nothing changed in transit.
d. An entry is made in the Electronic Record table for each file, where the checksum is immediately recorded.
e. An entry in the Object Group table is added to reflect each folder represented in the ingest batch. In the event that what is being ingested consists of shapefiles that follow standard naming conventions (same file name for each piece with expected extensions), then object groups called subgroups will be added per shapefile. More than one shapefile can then be contained in a folder but described individually.
f. Metadata extractors are run and write their data to fields in the Electronic Record table. The original XML file produced by the extractors is saved on the server and linked to the Electronic Record.
g. Descriptive metadata is extracted from FGDC files and written to the Object Group table. The original FGDC source is saved on the server and linked to the Object Group record.
h. A report is generated displaying the Electronic Records and Object Groups that were added, with a notation of which metadata extractors were run and any errors encountered.

Once ingested, then the user is free to add additional descriptive details, create their own subgroups, or rerun metadata extractors. The interface allows users to download whole object groups as a zipped file, or view individual files by bringing them back to the desktop and auto-launching whatever viewer the user has associated with that file type.

In Utah's testing of its data ingest feature, about 3,000 photos of 2 MB each were ingested successfully into AXAEM by pointing the client to the parent folder on a local hard drive. Since the ingest started at the desktop, necessitating the transfer of the files over a wide area network to reach the server, plus required FITS to be run against each file, this particular ingest took four hours for the system to complete. If the ingest had been run from the server instead of the desktop, the time required to complete the process would have been greatly reduced. Other data ingest tests took larger files of about 5 GB each, which also successfully ingested—the only constraint being network capacity to upload the files.

Figures 17 through 20 below display how the software looks while interacting with Electronic Records and Object Groups.

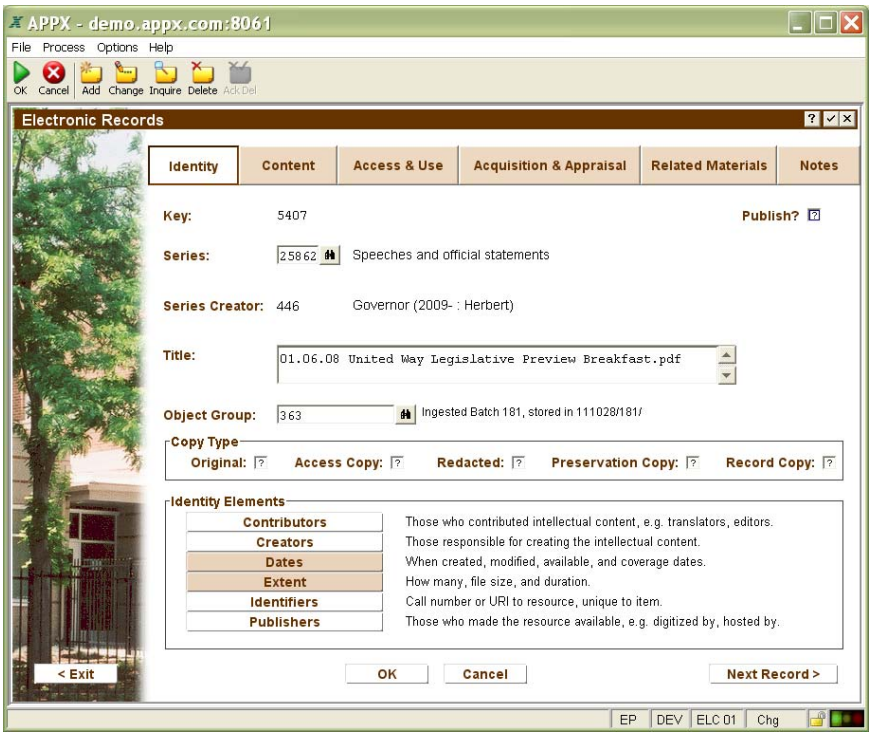


Figure 17: Example of an Electronic Record that was ingested

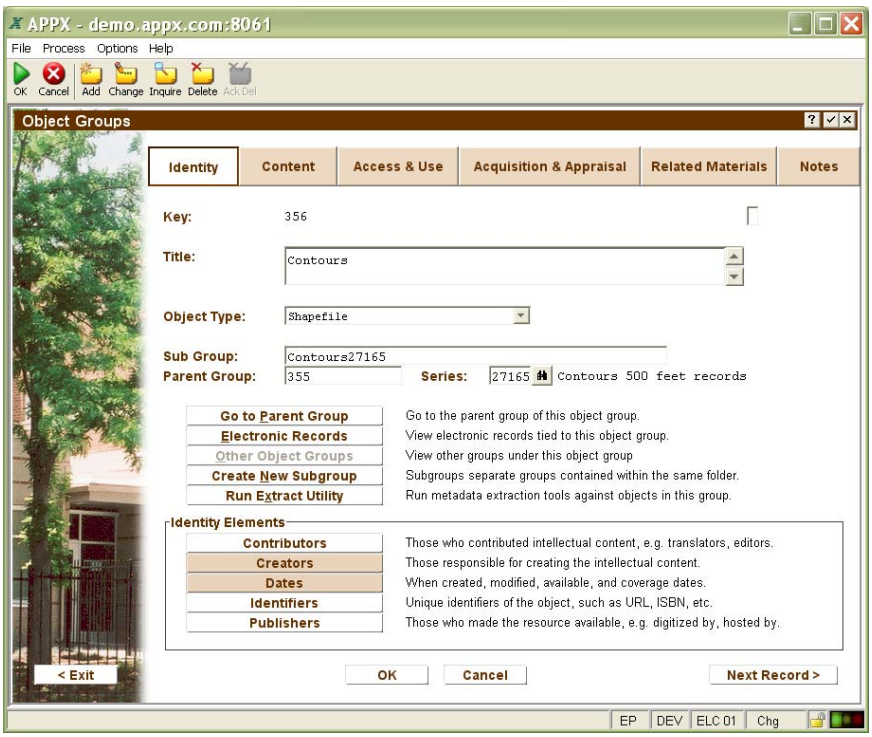


Figure 18: Example of an Object Group representing a shapefile

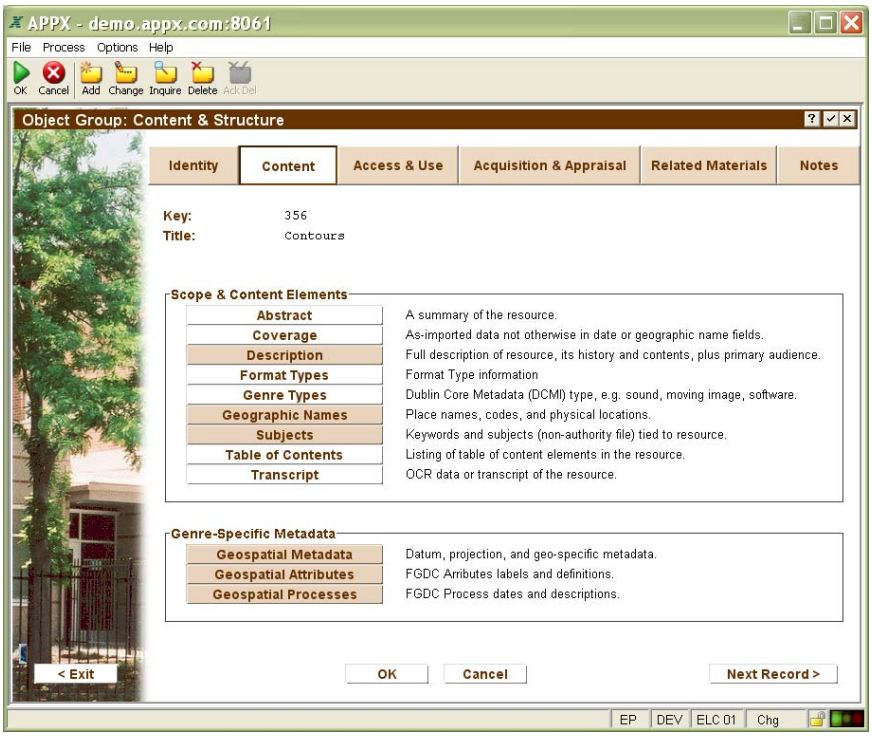


Figure 19: Metadata from the Content tab of the Object Group

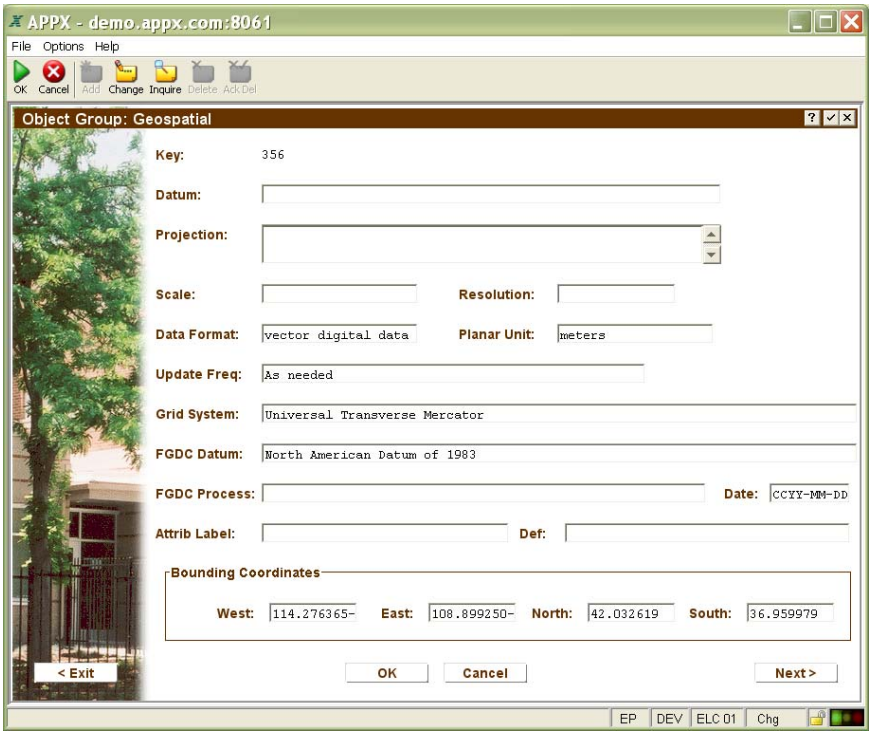


Figure 20: Geospatial metadata as ingested from the FGDC source file

To export metadata for reuse in other systems or to assist with data entry by moving data into a spreadsheet and then re-importing from the spreadsheet, options are available on the Electronic Records menu. To export, choose which fields you need:

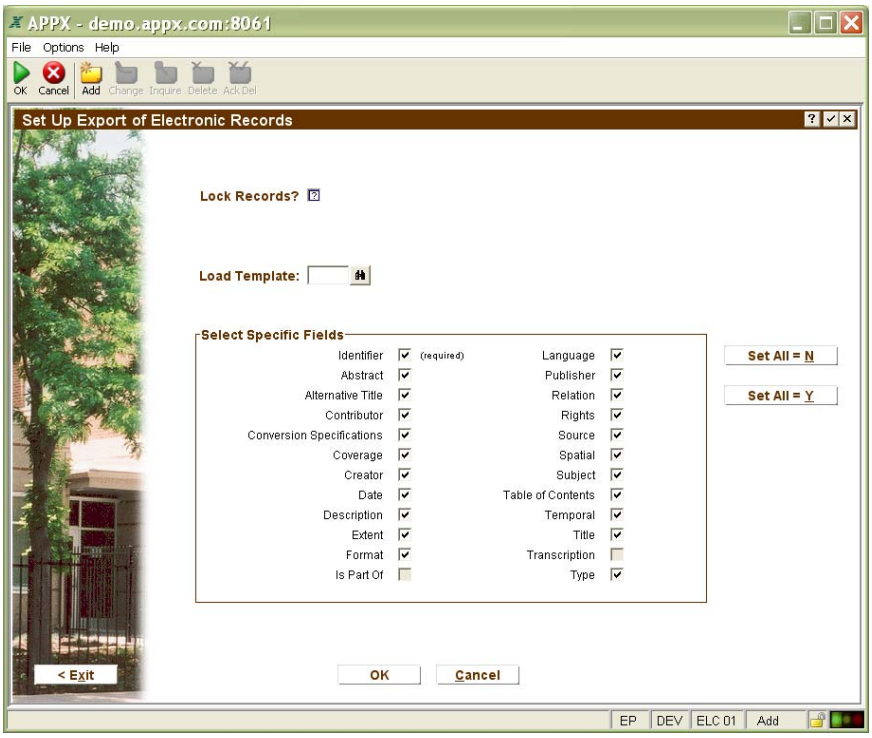


Figure 21: Electronic Record Metadata Export screen

The export will bring up the data in .csv form as a spreadsheet, where the user may use the auto-fill features of the spreadsheet to ease data entry. To import the records back to the system, select that option on the menu.

Media types are tracked by AXAEM, including individual media, so that an institution may manage migration schedules and other conservation needs. See Figures 22 and 23.

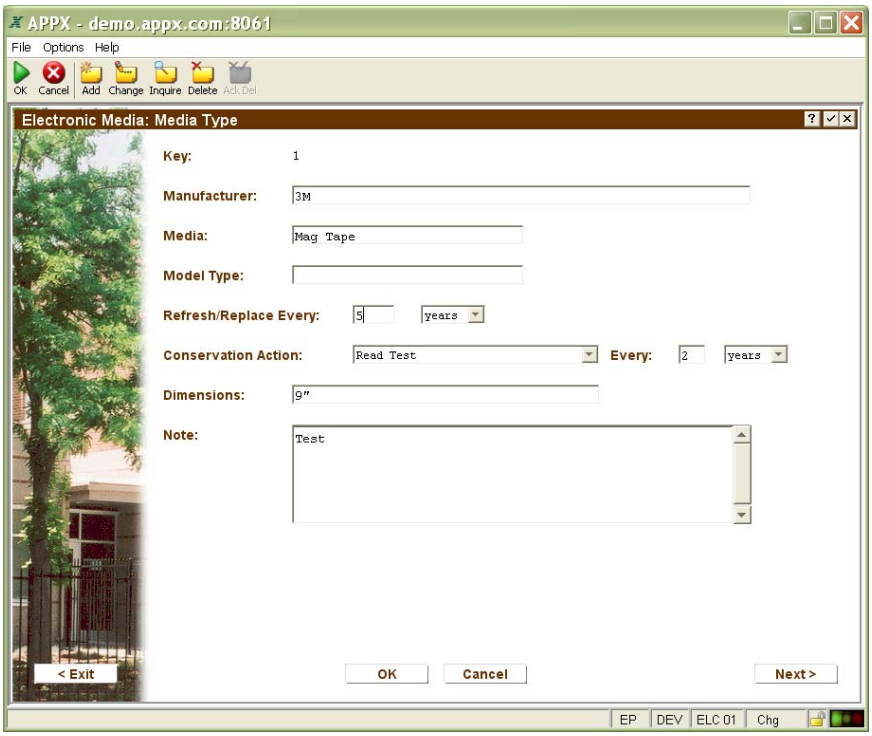


Figure 22: Media type description screen

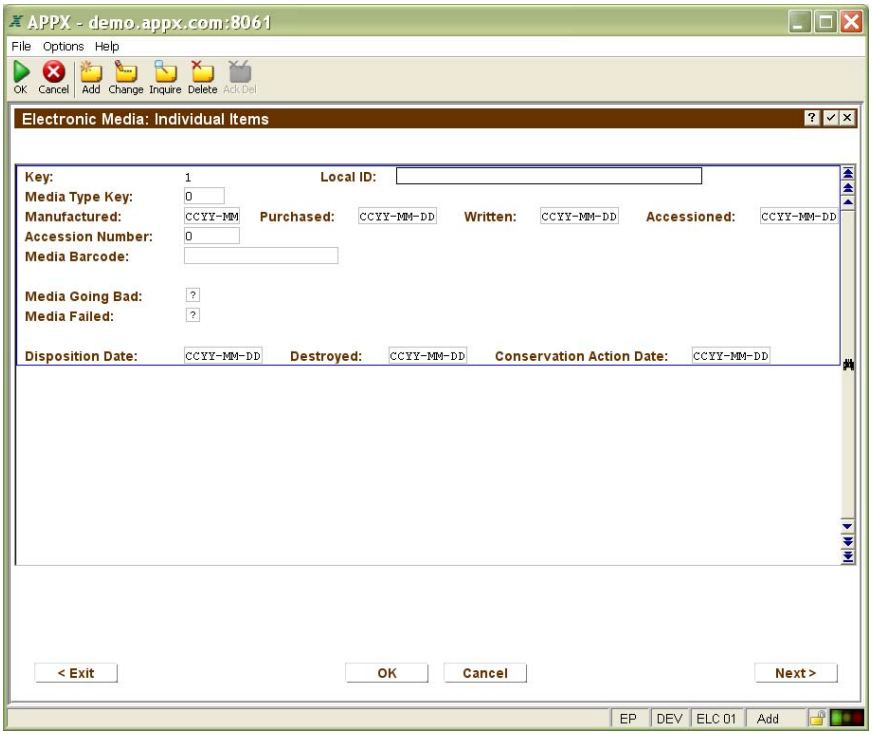


Figure 23: Individual media tied to a media type

Future Development Plans

A key long-term archival management function is the ability to transform records from a potentially nonsustainable format to a more contemporary, or better supported format. The AXAEM system's metadata extractors, run during ingest, attempt to identify the file formats for the records, and these file types are recorded in electronic record's metadata. This file type metadata can be used to facilitate the file type normalization or transformation process. New features are being developed to support these types of workflows and to develop transformation rules.

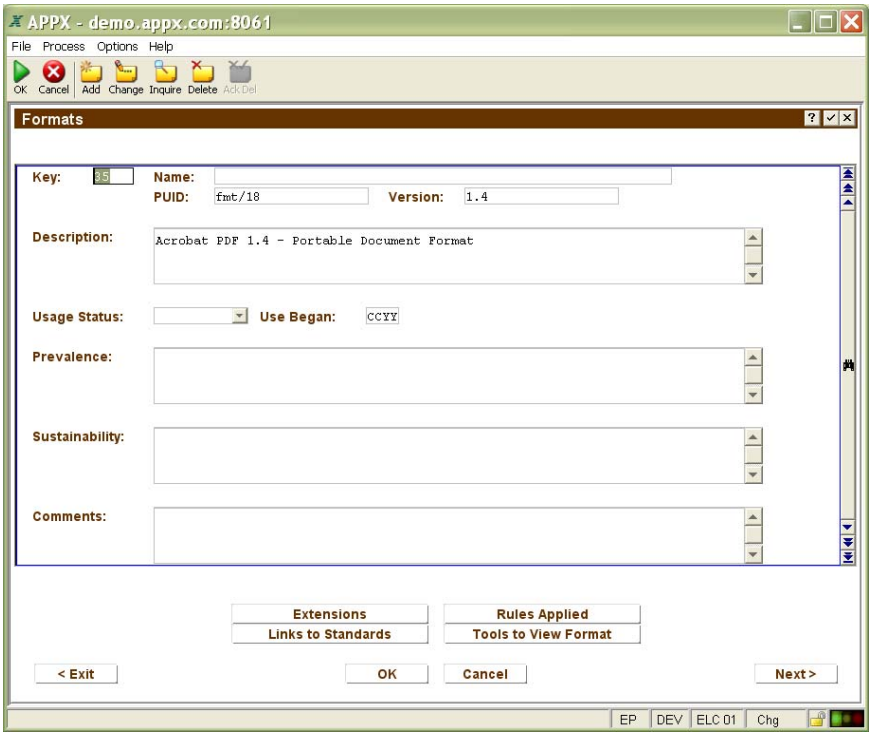


Figure 24: File Formats screen

The File Formats screen in Figure 24 already allows users to add details describing any format, whether that format is formally defined within the PRONOM[13] database or just locally defined. Functionality is also planned (see Figure 25) to include the ability to create "rules" that define system handlers, or automated responses, when a particular file format is encountered, for example, to transform one file format to another the archives thinks is more sustainable and prefers for long-term preservation.

[13] See http://www.nationalarchives.gov.uk/PRONOM/Default.aspx.

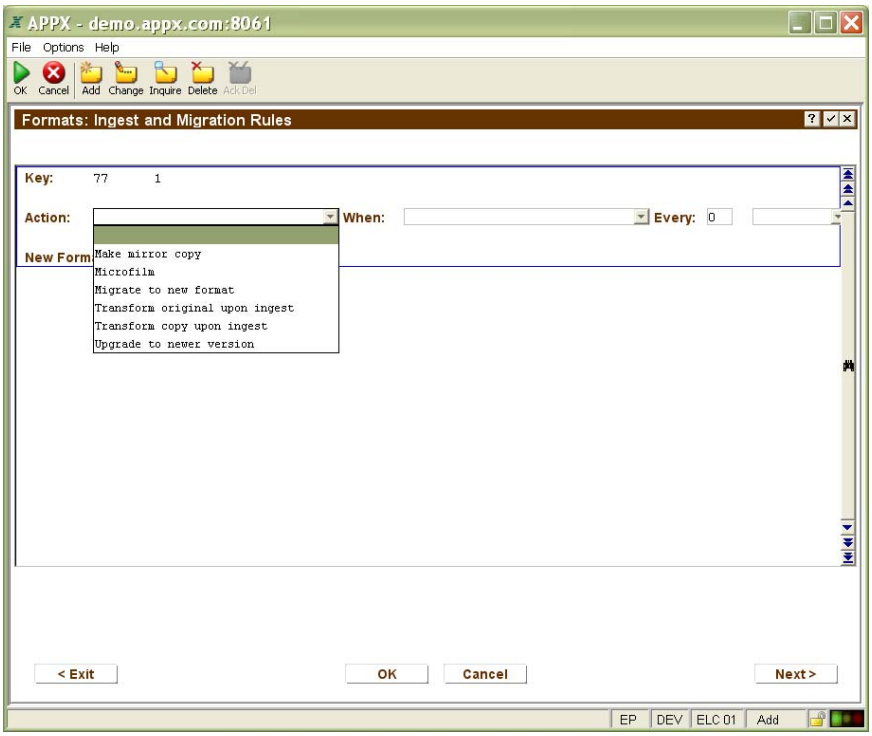


Figure 25: Format migration rules screen

The intent is that the system will allow the user to make multiple copies of any ingested record and specify where each copy will go. At the moment, the record is placed in a designated storage space on the application server. The user may tell the system where other copies are located, including hosted locations, but AXAEM does not currently move the data there automatically. In the future it will have this capability.

Small refinements to the data ingest process will continue to be made, such as including virus checking, ingesting data packages such as BagIt bags, .zip files, and others; supporting multiple types of checksum methods; and making more direct use of other metadata extractors such as Droid. More types of descriptive metadata could be supplied with submitted records, and more metadata standards will be supported to auto-map to AXAEM fields. Formal data validation will come with migration and transformation features, but may not have all possible features at the outset. Certainly more integration will happen between the electronic records part of AXAEM and its records management features such as retention schedules, plus archival findings aids.

The other major functionality being added to the system is the ability to search all of the metadata from the electronic file and object group tables (see Figure 26). This is being done by integrating the web search engine Solr[14] with AXAEM. Solr offers the web user faceted searching and very quick response, with results ranked by relevance or other desired value. The entire AXAEM database can be indexed by Solr, as well as data outside the database. Text-based files that are ingested may have their contents indexed,

[14] See http://lucene.apache.org/solr/.

too. The question then becomes whether the archivist wants the records indexed immediately, or should the system wait until appropriate access restrictions have been applied. Utah has opted to use a Publish checkbox on the front screen of the Electronic Record to indicate whether or not the record and its metadata should be made available to the Solr search engine. The ability for the user to modify the Publish checkbox field will be controlled by security rules that check the public access classification of the series and the item, as well as user security. In this way, only records which are classified Public, or which have been marked as a Redacted copy could be made available to the search engine.

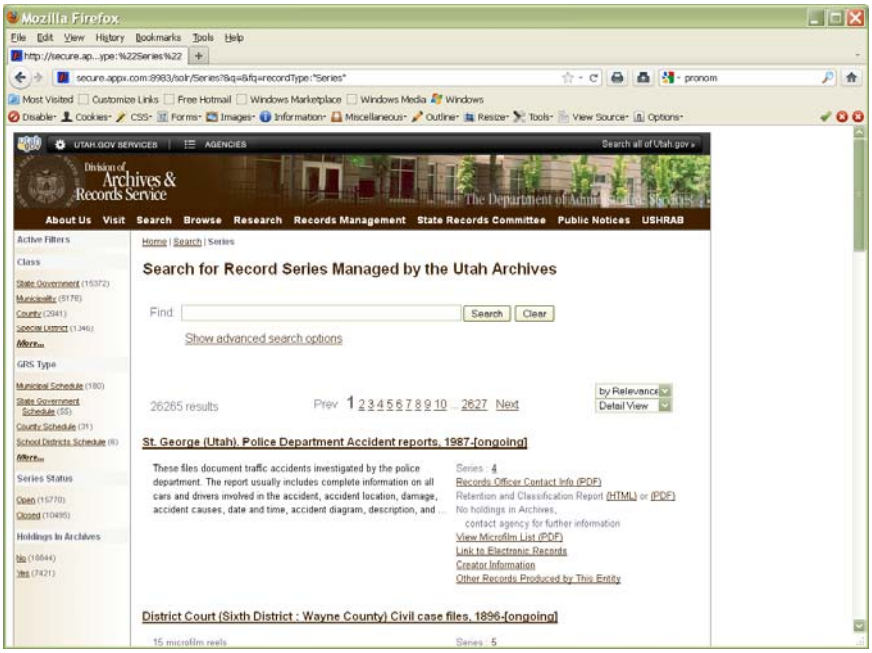


Figure 26: Solr search engine results for AXAEM bibliographic records

These new features under development, along with the comprehensive source-code review and update being done by the vendor, resulted in a system being evaluated that was essentially a beta release, with all of the attendant runtime bugs that rapid change engenders.

Evaluation

Testing was done twice. Initial tests were run during the spring and summer of 2010, and follow-up tests in the fall of 2011. Early testers were not given specific testing instructions (other than brief system documentation), but later testers were (see Appendix 2). Early testing results showed both promise and problems. North Carolina and Kentucky both liked the breadth of features and could see utility in managing an archives using a single lifecycle-based system, as well as using just a part of the system for one or two needed functions. The problems encountered by early testers fell into three general categories: the test environment did not have all parameter files and default fields in place yet, resulting in runtime errors that did not exist in Utah's production environment; limited documentation of features or understanding of how general APPX navigation works; and difficulties reconciling Utah's procedures and related system efficiencies to another organization that uses different procedures.

Later testing showed that many of the initial issues had been resolved. The runtime errors were largely eliminated (although still fluid due to the system being in development), and additional documentation was provided to testers. Also, feedback about how to make the system easier to navigate resulted in new features being added which strengthened the overall usability of the system. This included a redesign of the menus, data entry wizards being added, and a change in how scrolling lists display in the client. During the course of the grant, so many new features were added that both states saw the adaptability of AXAEM. This served to lessen the concern regarding state-specific procedures and workflow, although not entirely eliminate them.

The electronic records portion of the system was added between the first and second testing phases, so later testers could test things related specifically to ingesting geospatial records into the system, while early testers focused more on the retention schedule and agency history portion of AXAEM.

North Carolina

North Carolina's initial comments were these:

"The AXAEM system, developed by APPX, with a great deal of input from the Utah State Archives is unique in a lot of ways. What primarily interested the North Carolina State Archives in the system is how it covered life-cycle management of records from the creation of a description of an agency, through the retention and disposition schedules for the records of that agency (to the series level), through cataloging the permanently valuable records, through access with patron registration.

"AXAEM begins with the input of information about an agency, whether there are any actual records for that agency or not. You can enter identifying information about the agency ranging from simply giving the title, date of creation, structure within state or local government to who the records officers are for that agency. You can add notes such

as an agency history, legal structure and mandates pertaining to that agency, etc. so that you end up with all the possible information you need about that agency before you ever enter a description of any of that agency's records. All of this information can then be fed into a MARC-formatted catalog record for the records of that agency. Even Archival Encoded Context or EAC has been incorporated into an agency record.

"The first thing the NCSA staff tried was going through the Manual that had been sent. Because most of the records had been stripped out of the NC model, there were times the NCSA staff tried to follow along, but there were no records in a particular part, so they really couldn't follow it. Whenever they tried to enter new information, they always got error messages."

At this point, North Carolina attempted to add new agency records but they were impeded by the combination of runtime errors and lack of training. After some trial and error, they were able to add an agency record. The difficulties encountered exposed a challenge for AXAEM designers. For data entry, agencies must be added first before the bibliographic data and associated retention schedules can be added, since the creating agency is a required field in the bibliographic system (the bibliographic record is not technically a child of the agency). Yet adding agencies is one of the more complicated parts of the system, because it requires organizing these entities (corporate, personal, or family names) into hierarchical groups, and naming them according to established cataloging standards, including the use of MARC coding within the name. Once agency hierarchy is in place, it serves a vital function in the system to be able to relate one entity to another and pull reports and statistics based upon that hierarchy. Utah's current solution for managing agency hierarchy is to assign alphanumeric codes to each hierarchical level. Entities may have up to five hierarchical levels. The code for the top-level department must be unique, while the codes for subunits may be interchanged with other entities as long as the whole set of codes for a given entity in the hierarchy family is unique. This means that before users can add an agency, they must first establish a hierarchy code and understand its purpose.

System documentation[15] on Utah's website was available to the initial reviewers regarding agency hierarchy, although it is unclear how much that documentation was used. Documentation included written instructions as well as a Flash movie that walked the user through the process.

North Carolina explained their early struggles:

"The NCSA staff then tried to enter information about particular state government entities: The State Archives and the Center for Geographic Information Analysis. The NCSA staff entered Entity Structure Codes for:

1	DCR	Dept. of Cultural Resources
2	HR	Division of Historical Resources
3	ARS	Archives and Records Section
4	IBM	Information Management Branch

[15] See http://archives.utah.gov/axaem/axaem.html.

Then they went to Entity Input and tried to put in further information about each level. Clicking on Inquiry Mode, a list of the entities created already came up, but none of them were the four listed above. So they tried to re-create them. It was not easy. They kept being told either the agency was already on file, or the hierarchy wasn't correct. Finally they got everything entered. Then they tried to enter the address, but the city wasn't on file. So they went to the City/Zip Code Maintenance and tried to enter Raleigh, but the County wasn't there, so they couldn't save the record."

When Utah saw this response, they realized that the procedure of first adding hierarchy structure codes, and then adding agencies based upon those codes was very unclear. The problem appeared to be a misunderstanding about the two-step process. The name/title that is used to distinguish one code from another during the first part of the process may reflect the agency name, but the user may not realize that he or she has not actually added the agency yet, just the structural codes. After the codes are in place, the user adds the agency record and applies the codes to the new agency, then adds details about the agency, including the official name. Only until this part of the process is completed does the user have an agency record that can then be used to create the bibliographic record. Utah's other realization from these comments were that some default data needed to be put in place in the test environment so that testers wouldn't trip over missing fields. This data was added.

To ease the experience for adding agency information to AXAEM, a new data entry wizard was added. While the entity wizard does prompt the user for necessary details, parts are still complicated and will likely be redesigned. Eventually, the hierarchy codes may go away and hierarchy could be managed through some other means, but currently that is not the case. A separate wizard for creating bibliographic records was also added, which makes adding those records quite easy. After bibliographic records are added (in whatever order the user chooses), then they can be associated into a collection hierarchy that reflects record group, series, and subseries relationships.

North Carolina tested other features of the system, saw things they liked, but also ran into issues primarily with unfamiliarity about how the system worked and a lack of data to effectively run processes such as a full catalog export (the export process first checks the holdings status of the record prior to export, but box and microfilm data had not been entered into the test environment, so there were no holdings). Sometimes one user had difficulties where another succeeded, which may have reflected use of available documentation. When adding bibliographic records, comments ranged from "The NCSA staff could not figure out what to do about a records retention schedule that was neither an electronic record or a general schedule" to "You will be happy to know I added a record to the Office of the Governor and related a collection to the Military Correspondence. This was easy." Despite these problems, North Carolina's conclusions from their initial testing resulted in these comments:

"The system definitely has potential. As mentioned at the start of this evaluation, the system is very appealing because of the potential to record everything from an agency's history and organizational structure to managing that agency's archival records, to

assisting researchers in gaining access to government records. However, the system is not very intuitive and needs a very detailed manual to accompany the various functions of the system. It would also be useful to use more standard terminology for archival material such as “agency” instead of “entity”; “agency abbreviation” for “structure codes”, etc.

“In addition, while testing the system, it was changed and other features added, which made it more confusing to test. It would perhaps be better to evaluate this system once it is essentially completed, along with a user manual, rather than while it is still under considerable development.”

During later testing, however, North Carolina made these comments about the system:

“On the whole, it is much improved from when I looked at it last spring. It is easier to intuit how to use the system and your help screens were a tremendous help in moving through the processes” and “I do like the system a lot more now” and “I really, really wish I had more time to look at the system.”

After ingesting geospatial records, another North Carolina reviewer said, “In general, I can see the potential for AXAEM as a preservation platform ... with the caveat that to pursue a real implementation would require more formally defining our organization’s particular end-to-end workflow, and investment in the associated custom development/customizations/configurations and refinement based on the particular of how NC State Archives would want to store and “tag” its geospatial data collection.”

Later testers also produced some specific feedback about how to improve AXAEM even more. “I had a chance to do a test upload of a dataset and browse around the interface a bit. It looks to be on the right track – with regard to the archivist’s functional tasks that are supported through the user interface. I was able to successfully upload a dataset, and the FGDC metadata was extracted to populate the AXAEM metadata fields. [See] Object Group: 362, Electronic Records 5400-5406. Most of my feedback is offering some observations and suggestions with regard to the user-interaction / user-experience.”

The detailed feedback provided by North Carolina reviewers and Utah’s response to issues and workflow questions can be found in Appendix 1.

Kentucky

Kentucky also tested AXAEM. After they completed their early testing, they shared these comments:

“The Kentucky Department for Libraries and Archives (KDLA) agreed to examine APPX as part of the GeoMAPP project. KDLA uses Dspace, an open source data repository software, to store, describe, and provide access to a variety of digital objects that it has accessioned into its archive. Additionally, KDLA uses Microsoft Access to create records retention schedules and track records accessions.

"One advantage to the application created by Utah using APPX software is that descriptions of both paper and electronic records, as well as records retention schedules, are all integrated into a single system using the developmental capabilities of the software. I can think of few archival software packages that attempt this kind of integration. Additionally, the adoption of standards such as DACS, EAD, MARC, and Dublin Core, as well as the use of FDGC metadata elements are unique and attractive features of this application. The automated import of descriptive and technical information should be a useful tool, if it can be done as a batch operation that does not require technical intervention in order to integrate the electronic record and the extracted metadata.

"While Kentucky's testing of the system was less involved than North Carolina, the value of the generic application was apparent enough to Kentucky to recommend it to archives and records management units, particularly ones with data repositories, to investigate and examine its adaptability to their environment. Unfortunately, Kentucky was unable to devote the time to establish all the elements needed to integrate its current systems with the Utah APPX application."

Utah

Utah's experience with this product over many years shows that AXAEM is not just helpful for managing their archival and records management responsibilities, but vital. It is a workhorse that allows Utah to operate highly efficiently. With this system, Utah is able to produce output comparable to institutions that have many times more than their 25 employees. In addition, since the product has been in use for so many years, its adaptability has been put to the test quite frequently. Utah has changed operating systems, servers, and databases more than a few times, and application changes are part of the daily work process. These changes have taken place smoothly, especially since the application itself does not need to be changed when the underlying OS, hardware, or database changes. System downtime is generally quite minimal. In short, AXAEM is robust.

Utah's investment alone in the system might keep them continually using it, but it is also true that if faced with other software choices, AXAEM still measures up and offers options that are hard (or impossible) to find elsewhere. Chief among those options is the reality that AXAEM changes with you. Regardless of what functions are currently in place, the knowledge that more can be added or changed as needed at minimal cost keeps the system continually in use. Utah archivists are simply too used to being able to make the system do whatever they want to easily give that up, and nothing else automates their archival workflow to the degree that AXAEM does.

With that said, however, the system isn't perfect. The ease of use could be improved, and will be. Ideas that have come out of this evaluation process have not yet been implemented, but should improve many things greatly. Some of the features that have

been put in place have been developed to a degree which is acceptable to Utah, but may fall short of expectations for other institutions. Without knowing precisely what those expectations are, however, Utah hopes that others will see the good that the system offers and will want to add upon it to the benefit of all.

Conclusion

While Utah obviously has a bias towards this system and values it not just for current features but also expected future capabilities, North Carolina and Kentucky evaluated AXAEM based upon the functionality already offered, without any real knowledge of how to change something they didn't like. Both states saw its potential and reacted favorably in their final conclusions. Kentucky shared this with their recommendation, "You have created an attractive system that is feature rich." Both also realized that for AXAEM to assist the workflow in their environments, customization was likely needed. Still, that customization should not be a stumbling block, and much could, in fact, be done by a tech-savvy archivist trained in the application design features of the product rather than a full-time programmer.

While much functionality has been added to AXAEM, there is still much left to be done. During the GeoMAPP grant project, participants were exposed to many products and learned about particular system requirements desirable for managing electronic records. These ideas served to provide Utah with a roadmap for future system development, both near-term and long-term.

Appendix 1: North Carolina and Utah Email Exchanges During Testing

North Carolina's observations are numbered below, followed by Utah's response, structured mostly as the conversations appeared in the back-and-forth emails:

Issue 1

As I navigated through the interface, I guess my major feedback is that it would be nice to have some sort of contextual location visual indicator, especially given the breadth and depth of the options. You kind of lose the sense of where you are in the overall navigation as you navigate down into the data entry screens.

e.g.
Basic Menu: Search & Edit ->
Search & Edit -> Electronic Records Menu
Electronic Records Menu -> Electronic Object Groups
Object Groups ... the windows tell you you're in the "Object Group" and which tab but you lose your overall sense of location in the entire hierarchy (got especially deep as I looked at Electronic Records within an Object Group).

Utah's response:

We should be able to add more breadcrumbs to the detail screens, although it's possible to dive into electronic records from more than one place, such as if you are looking at the bibliographic record (series), and want to see electronic records attached. That may complicate how we do the breadcrumbs. The more you use the system, the more you get a feel for where you are, but we could help the user out more.

One idea suggested to the vendor was to add a pop-up box that keeps track of screens the user navigated to get to where they currently are, and be able to click on links there to navigate back. While such a feature is not in place yet, the indications are that such as thing could be done.

Issue 2

The [<= Main Menu] link at the bottom returns you up one level, until you return to the first "Basic Main Menu", though there is really only 1 Main Menu for the application. Would be helpful if the label reflected the name of the menu it is returning you to.

Utah's response:

At the moment I'm reusing the subject-based menus in more than one place, although if people decide they like the new menu design, then I'll make sure all buttons are properly pointing back to the previous menu. So in other words I haven't changed the label on the button that says <= Main Menu, even though doing so would only take a second. Do you guys like the menu organization, or would just seeing a list of subject-based menus make more sense?

The issue here is which set of menus works best for users: one that offers a list of subject categories (agencies, electronic records, general schedules, microfilm, etc) on the main menu, or one geared toward new users based upon actions (add & import, search & edit, print & export). The latter menu system walks the users through the processes and explains more about what they are for. The subject-based menus get the user to the place they want to be more quickly. From other feedback received by Utah, the menus should probably reflect actions rather than subjects, although quick links to subject-based processes could be made available in dropdown lists from the toolbar

Issue 3

The scrolled list windows are a challenge. When you first enter the screen, you do not see the slider bar, so don't immediately know that it's a scrolling list. When you mouse over the list section the scroll bar appears. However, the scrollbar is always positioned in the middle of the vertical space, and gives no visual indicator of your relative location in the list. Some UI widgets -- the scrollbar size itself will also auto-size to reflect the % of the entire list that you are viewing. Then when I scrolled down to the bottom of the list, the scrolled window always "got lost" and there was no way to re-display the scrolled list.

Utah's response:

I agree that the default scrollbars as APPX defines them are essentially useless. From a software point of view, the difficulty is two-fold: knowing the number of records in a table and especially the number being constrained to a particular value (which changes on the fly and is harder to store), and being in a client/server architecture. The latter means that the client doesn't know how many records are on the server, and if the server sends all the records down to the client, there could be performance issues depending on the size of the file. APPX can count records fairly quickly, but if a file has billions of records, the vendor is hesitant to always make it count, so they have been debating among themselves the best way to tackle the problem. A better model than having a scrollbar is to use page views, similar to how Google searches divide up a list of results. I think down the road the screens will look much more like that, and the scrollbar will likely disappear.

You may use your keyboard's Page Up/Page Down keys to scroll through records. To get down to a record that's way down the list, go into Key Entry mode, meaning click either

the Inquire Mode or Change Mode buttons until it provides a blank screen with an open key field (or, if scanning, use the supplied search field to enter a value). You can change the key path by then hitting F3 and selecting a different one. The vendor agreed to change the client settings on your webstart instance to display a different kind of scrollbar that lets you click buttons to page up/down or jump to the top or bottom of the list and anchor it to the screen so you can see it immediately. This other scrollbar should now be in place (it has always been available to the client and could be seen by use of settings, but it was not the default scrollbar style). You may also scroll down using the arrow keys if your APPX client has that setting implemented.

The scrolling window could be any size from a single line to a whole window. To see everything, just use Page Down until you can see that there are no more and/or a message appears at the bottom saying “No More Records.” As I will explain further down, the Next or Next Record button simply takes you to the next item in the table, so if the window is full-size (which happens when there are a lot of fields tied to the record), then the whole screen refreshes, and if it is a single row (or a 5-row window, for fewer fields), then only the next row is highlighted.

Issue 4

Identifying Object Group - by Objgroup key. When this ramps up, I think that you'll want to have some other mechanism (e.g. search by string in Object Group Title) to identify the Object Group other than its numerical identifier in the database (or search by string in Electronic Record Title to identify an individual electronic record).

Utah’s response:

I’ve added a feature to the Electronic Records menu, Search/Print tab that lets you search either electronic records or objects groups by title or by series/bibliographic number.

Issue 5

Electronic Records Section

Searching by Electronic Key -- users won’t necessarily know the Electronic Key ... and browsing through 100s (?1000s?) of individual electronic records will be quite unwieldy.

I also experienced issues with the scrolled window here as well. As I tried to advance through the scrolled list by pressing down the down-scroll arrow - the scroll list would “get lost” (as above) I could never “fast” scroll down to the list to see my record 5400, if I carefully single-clicked the down-scroll arrow, I could advance through the list.

In general -- there is a [Next Record] button ... it would be nice to be able to also access [Previous Record]. Once you have a 'detail' page, you cannot select another record, you must [Cancel] out, and re-enter the Electronic Records form to select another record.

It would be nice if there was a way from the Electronic Records (individual item) view to display the other members of the same Object Group -- especially as you have the option of reassigning the electronic record to another object group. (You can view the object group members from the Object Groups screen)

Utah's response:

See above. You can jump down to a specific record without scrolling through all records. I think you'll find that when looking at electronic records, you will most often want to interact with them in context of their bibliographic record, so you wouldn't need to see the whole electronic record table all at once. However, even with that, the number of electronic records or object groups tied to a series could be lengthy, so being able to sort them from within the series screen will be important.

If you are trying to scroll through a list of records to decide what to pick, I'd use the "Search by" screens I just added rather than the scrolling lists you get from clicking on the Scan button (the one with the binoculars). It's easier to go back and forth and pick what you want.

The "Next" and "Next Record" buttons are doing identical things. They will take you to the next record in the current list of records that you are looking at. For row-by-row scrolling lists, this is kind of redundant, since it's easier to just click on what you want. For full-screen scrolling records, it's probably clearer to click "Next." The existence of the Next button tells you that there's more data in the file to look at, but not if the screen is scrolling or not. There are advantages and disadvantages to making a full-screen record scrolling. The tabbed input pages that you see in Bibliographic Input, Agency Input, Electronic Records, and Object Groups used to be scrolling records, meaning that when you clicked OK it would automatically advance to the next record. We changed them to be non-scrolling, so that clicking OK would take the user back to Key Entry mode (a blank screen). This helped users understand that they were no longer in the record they had just been editing (problematic when they intended to go to another tab, not another record). One of the effects of making a screen scrolling or not is the ability to navigate back up to the previous record. On scrolling screens you can go back up to the point you started (not necessarily the first record in the file if you initially used Key Entry to start somewhere in the middle of the list). Non-scrolling screens will not page up.

I do think we need to change the navigational buttons at the bottom to be a little more context-sensitive. Mostly I just use the same design template that has all these same buttons, so they appear places where it might be better to change how they are labeled. I think you will see a lot of changes here before AXAEM gets released.

I agree with the idea of being able to see object group items while looking at electronic records. As we start using the electronic records features more in my office, I expect a bunch of new usability requests to come my way. I've been gathering a list from people. One suggestion is to make the object group title reflect the name of the folder that it represents, with qualifications from subgroup names as needed. Another idea is to create electronic record subgroups (just as object groups have subgroups when more than one multi-item object is housed in a single folder). In the e-record subgroup, however, it would give you the option of describing a part of a record, where one single physical file has many descriptive items in it. One example would be clips from a movie.

Issue 6

FGDC Extraction – abstract

The <abstract> and <purpose> fields are both loaded into the [Description] field.
And AXAEM does include an [Abstract] field.

Utah's response:

I deliberately concatenated the contents of the FGDC <abstract> and <purpose> fields to the Dublin Core <description> field since that seems appropriate for how Dublin Core defines description. I think of an abstract as a more narrow description, so it has a different purpose. I could make the system copy <abstract> data to both fields. Alternatively, users could re-map the metadata to fields they prefer. That would require some programming knowledge.

Issue 7

Conducting a 2nd import on an existing Object Group I noticed if I did a 2nd import -- the metadata from the new import was added to the record -- creating a duplication of metadata records.

Utah's response:

We were debating what behavior the system should have when updating metadata, whether we wanted to keep the old or not. On the safe side, I said keep the old. I'm guessing you would want to have the option to just replace, not append? Probably in the future this will be handled by system preferences that users could set.

Issue 8

Object Group => Electronic Records

Object Group “management”

I noticed that there was no way to remove an item from an Object Group - in the Object Group interface -- For example, being able to manage the Object Group composition from the Electronic Records List in the Object Group. I also noticed that if I selected the [Detail] of an item, I would navigate to a “read-only” view of that individual record. [OK] returned me to the Object Group list (see above) However, the [Cancel] kicked me back up to the Object Group page ... seems like the [Cancel] button is irrelevant on this page, as there are no modifications a user can do. Consider using a button labeled [Return] rather than [Ok] and the [Next Record] was non-functional though it looked “active” -- here it might be nice to advance through the record detail screens within the Object Group, and then disable (and dim) the [Next Record] when you’re viewing the last record in the Object Group.

Is it envisioned that you might remove an object from an object group on the “Electronic Records” screen by deleting the “Object Group” field value? That’s similar to the individual file access that’s offered through the Electronic Records interface -- for example, to be able to see the individual file size (Extent) or Dates.

Note: selecting the [Next Record] button here -- advances to the next Electronic Record in sequential order ... refreshes the entire screen with the next Electronic Record (compared to the [Next] button above that changes the mouse / input focus to the next item in the “list” views - such as the list of electronic records in the object, or the list of subjects, or the list of attributes, etc.), and, as it does in with the [Next Record] button in any of the tabs (e.g. [Content], [Access & Use], etc.).

Utah’s response:

I think it would be good to be able to manage objects in the Object Group screen. I’d also like to be able to delete whole object groups and have the system also delete related electronic records and remove associated ingest files from the server. Other times, I might just want to disassociate all the electronic records with the object group but keep the e-records and ingested files intact. I imagine that down the road we should be able to do these things. For now, you can remove the electronic record from the object group only by editing the field in the Electronic Record window, and I’d do it by first navigating within Electronic Records instead of while looking at them after first drilling down from the Object Group. I think you can go into Change Mode after drilling down, but the constraint to the object group will limit what you can change, and may produce unexpected results or an error that you wouldn’t otherwise get when going in the other way.

Other Comments

One of North Carolina's later reviewers spent more time interacting with other features of AXAEM. Here are those comments:

The first thing I did was to try again to create a new bib record, using the "Add Bib Records Wizard." The local key number appeared to be the local call number...is that right? How would you handle a series of records? For example, if I wanted to describe a group of boxes of probated wills, would I have to describe each one or is there a way to do them as a group? The local bib ID is limited to 14 characters, so I could not enter the full call number describing the full run of boxes. In addition, I had made a note that the scrolling wasn't working when you chose the scan button, which I believe you said had been fixed. Another question I had made note of at what point do you add a new creator ID? Do you have to know to do this before you ever enter a Bib record? Or is there a way to do it while in the wizard? And then, I had a question about access points. Since the MARC coding needed is between 600-799, can you restrict the list of codes that appear under the scan button? And then, finally, why is it that only dates appear under the scan button used to "assign subjects"?

In regards to importing records, I got an error report when I tried to use the access point wizard, under creating the bibliographic record; I could not add a new access point; and I could not add subject terms. Of course, this may be because of not understanding something about the system.

When I went to import the digital records I realized that I would have had to previously put in the information regarding the series, separate from the creator, to be able to locate the series to be used on generated records. Luckily (since I was importing some 1884 fair exposition pictures) there was one called "State Fair payments Exposition Division" which I used. If I had had transfer or accession documentation, would it have needed to be in a particular format? I did have an index to the images...could there be something I could have done with that during the import process?

Utah's response:

Your comments help me understand more about what you are trying to do. I didn't realize you would be attempting to add subject access points. That's an advanced topic! Thanks for letting me know that the access point wizard was broken. I wasn't aware. When you hit the Scan button, it displays a list of values in key order. In the case of subject headings, it's alphabetical, which means the numbers in date spans get listed first. You can always go into Key Entry mode while scanning to get down to a point in the list that you want to be, such as subjects that start with R. Subjects in the system are really segments of subject headings, not the full access points (e.g. Flower gardens--1860-1910). In the example, there would be one entry in the subject list for Flower gardens, and another for 1860-1910. The wizard is intended to help walk you through adding each piece so you end up with the full access point.

What length do you need the local key to be in the bibliographic record? Changing the field size is easy.

Your question about the boxes reflects a different approach to handling records than we use, which I find interesting from a system design point of view. In Utah, we generally do things in this order:

1. Create an entity record for a new agency, and associate it with its parent agency, if any. By now, almost all agencies that exist are already in our system, so it's rare for us to need to add one unless the government restructures itself.

2. Talk to records officers in the field to find out what kinds of records they have. Inventory the records to determine what the retention schedules should be. We follow the classic definition of a series as a set of records filed together that share a common purpose and retention.

3. Add a bibliographic record to the system which is tied to the agency key created above. I think it would be very nice to add the feature of being able to create the entity while in the bib-record wizard. The bibliographic record generally represents a series or collection as inventoried in the creating office. Part of the data in the bib record is the length of retention, as negotiated with the agency, as well as basic descriptive information to be able to identify the records. You can technically make any bib record reflect any level of description you want, including record group, sub-series, or even item, but we generally just use it for series. The hierarchical relationship between bib records is recorded in the Collection Management part of the system, but that would be added much later in the workflow.

4. The bib record then goes through an approval process to make the retention schedule legally binding. If the bib record reflects content described in a general schedule, then it is considered already approved. Once approved, agencies may interact with our services in regard to those records, and destroy them or transfer them to us if need be.

5. Boxes come in, and the transfer paperwork includes the unique ID (series number) that we established in step 3. We add each box to our Versatile Enterprise[16] system and tie it to the series number. Versatile then produces a barcode that gets put on the box, and it is associated with a location on the shelf. Versatile is linked to AXAEM in that it uses the retention schedules provided by AXAEM. AXAEM, on the other hand, displays boxes that have been added into Versatile. Generally, the only descriptive information we put on the box at this point are folder ranges and dates, although there is a field where you can add long descriptions. The software does have features to describe each folder or even each item inside the container, but that's usually too much data entry for us. Each box is automatically tied to the description in the bib record, so we know what they are. For permanent records, the contents will be re-boxed and described separately anyway during processing, so that's partly why we don't do it upfront. If you do not have a space

[16] See http://www.zasio.com/.

management system such as Versatile, then I believe AXAEM can use its older box location system. This hasn't been touched in some time and will need updating before AXAEM is released. I believe the plan is that it would be used by institutions that have no other software managing their records center or physical holdings, but is not ready to be interacted with yet.

6. Boxes stored in our records center are not yet in our custody. We run reports to see what has met retention and can be destroyed. Similar reports are run to see what has met retention and should be transferred to our permanent custody. A pick list is printed in shelf-location order and boxes are moved off shelves, headed to the shredder or our permanent records room. We don't run both types of reports at the same time and get the two sets mixed up. :-)

7. Once in the Archives, they sit for a while until a processor has time to work on the whole series. At this point, records are re-boxed, re-foldered, arranged, and described. Depending on how the finding aid needs to be structured, descriptions could be at the reel, box, folder, volume, or item level. These descriptions are recorded in the container list tied to the series. The processed container numbers that appear on the finding aid (e.g. Box 1, Box 2) are then linked to the Versatile box barcodes, so we know exactly how the archival intellectual information is tied to a physical medium, and its shelf location. To do this, we simply scan the box barcodes in the order that we want the processed box numbers to be, and the system adds it. Once a processed box number is in place, we can add other metadata to it, such as identify media that's in the box or attention that box items need, such as repairing, flattening, scanning, etc. We can run reports off of this table that will tell us all boxes that have VHS tapes or whatever in them in case we want to take some kind of bulk preservation action. It's a quick-and-dirty way to record what's going on with a box as it is either being processed or being used by patrons and Research Room staff who might want to make note that the box contains photos even if the series description never mentions them.

8. While processing, the option is available to organize discrete bib records into a hierarchical relationship. This is done in the Collection Management screen. First add the collection record, then tie the highest bib record to it. From there, you can drill down and add child records associated with other bib records. This means that you can process your records in any order without regard to hierarchy, then assign hierarchy later, or move things around easily if you find more records later that you didn't know you had.

9. If you want to assign local IDs to the bib records, that would probably come last, since they would reflect the final manuscript/collection/call number associated with the set of records, that would display on the finding aid. You may have other uses for the local ID, and that would be interesting for me to hear about so I can understand what your needs are.

10. After processing, then we tend to digitize the records and reuse the metadata contained in the bib record or the electronic record. If we want to index any particular series, we set up a table in the database that reflects the information contained in the

records. Data entry typically happens on a spreadsheet filled out by volunteers, who link the indexed data to the scanned file name. When this is finished, I import the spreadsheet data to AXAEM, which incorporates it into our online indexing system. The plan is that this process will not require programming for each series that gets indexed, which it does right now (not much, but some). The vendor will be addressing this need prior to AXAEM being released, because they realize it is important, and most institutions aren't going to want to do programming for each record set (personally, I like the flexibility, especially when I think about how we are going to ingest databases and make the content searchable). I don't know the details of how exactly it will be restructured, but you should still have the ability to add whatever set of fields you need to adequately index your records.

If your record indexes are in a form that would be akin to Dublin Core metadata, we are already working on being able to ingest that at the time the electronic records are being indexed. We've mapped the CONTENTdm Dublin Core metadata as exported by that system so that AXAEM understands where to put the content. It's assuming that the <identifier> field will be populated with the file name, so it can associate the metadata accurately with the record it is describing. Someday it will be probably easier to harvest directly from CONTENTdm, but we aren't there yet.

To answer your question about transfer documentation, it does not need to be in any particular format. All the system does it let you provide a pointer to where the documents are stored locally. If your desktop has access to that location, you can click the button and view the paperwork. We should probably update that so that you can actually ingest the documentation, so the pointer doesn't break down the line.

North Carolina's further response:

Yeah, I had tried once before to explore more than I was supposed to and was just seeing how things had improved! I do like the system a lot more now.

Once you mentioned that the subject headings, being alphabetically, started with dates, that made perfect sense to me...I've seen that in other systems and just didn't think about that possibility. That's what I get for being one of those people that just dive head first instead of reading instructions and/or asking questions. I like that I can go into Key Entry mode while scanning...that helps. I also like that the subjects are broken up (post-coordinate?); that is how we handled subjects in MARS.

The longest local key I can think of, for a series of boxes would be 29 spaces if periods count [C.R.100.508.1-C.R.100.508.300]. It would be nice if the system allowed the user to adjust if needed?

Steps 1 & 2. Maybe it is because N.C. records date so far back that we have different approaches. What you described in steps 1 and 2 is exactly the same sort of thing our records analysts do and enter into our SCHEDS system. But regrettably there is no

connection between the SCHEDS and MARS systems. So the archivist has to start from scratch in the MARS system when they start to arrange and describe a record. Plus, it may be 60 or more years from the time we receive a record to the time we actually get around to processing it! That's why it would really be nice to be able to create the entity while in the bib-record wizard.

Step 3. How much of the 'data entry' do the analysts do and at what point do the archivists start adding data?

Step 4. Is it your plan that eventually agencies can interact with the data base itself, adding records and their retention schedules or changing an existing one and putting it in the queue for approval? One of our dreams is to have a system where the interaction between the analyst and the agency, after the initial interview, is done within the database until the retention schedule is fully approved.

Step 5. So it sounds like you handle the retention schedules in AXAEM, but handle the space/location by Versatile? We are in the process of creating a barcoding system, so we are having to take our location system called FAIDS, which is somewhat integrated with our SCHEDS system, and break up the location ranges into individual box spaces so that we can link one space to one box (or a range of spaces to a specific range of boxes) via barcodes.

Step. 6. Do you run the reports for destruction, etc. within Versatile?

Step. 7. When the archivist processes a collection, do they do appraisal and remove any material? How do you handle changes to the Versatile box barcodes once a collection is processed? Or are there no barcodes in the Versatile system until the material is archived and processed?

Steps 8-10. So it sounds like the archivist is entering the bib information into Versatile? Or is it in AXAEM? I get a little confused as to just what system the archivist is in. Or are they having to duplicate the information that was originally entered into AXAEM by the analyst?

It sounds great that you are mapping to CONTENTdm...I will keep an eye out on your progress there. And, finally, it would be great to actually be able to ingest transfer documentation, even if it is just scanned pdfs of the original paperwork...

Thanks for working with me on this! I really, really wish I had more time to look at the system.

Utah's further response:

I have changed the length of the local key to be 30 characters. Someday we can show you how you can restructure the database on your own when you need to, it would just involve a bit of training.

We have really old records, too, that have not yet been processed. Everything does have a series ID associated with it, however, which we consider our most basic intellectual control. Not every series has a description yet (most do), but they all at least have a title. We consider title, dates, arrangement, description, retention, appraisal, and access classification (public, private, etc) required fields for all new entries into the system, and the bib record is always linked to a creating entity. For some records that have been around for a while, not all of those details are available. This is what the records analysts add when creating new retention schedules. If the records are already in our custody from years ago, then the processors add these details into AXAEM. Whatever data is in AXAEM about a given series, the processors take that and add upon it when they start a processing project. They can lengthen the description, add subject headings, or fix any number of details. You can look at our instructions about how we process containers at http://archives.utah.gov/axaem/Axaem-How-to-Process-Containers.pdf .

We have a web form that agencies can use to submit new retention schedules: http://archives.utah.gov/recordsmanagement/forms/inventoryform.htm . When the data is submitted, it goes directly into AXAEM and an email is sent to the records analysts. The analysts look over the submitted data, change anything that needs changing, click a button and it becomes a series. Once a series, then agencies can view the data from our website, and either print a report or send us more feedback about what they want edited (see form above under the heading Edit an Existing Schedule). Edits then get sent into AXAEM into the same holding area that original submissions go until the analyst looks things over. If you want, you could let agencies edit your database directly. You could give them access to the same webstart client that you are using and teach them how to add and edit. I would suggest that the data is presented to them in a way that limits their ability to view or change data they shouldn't, and also make it super user-friendly so they don't need much training. I'd probably just present them with a wizard and not let them navigate off of that single view. Your website can launch AXAEM and even constrain the screen to a particular record. The first time you click the link to the software it will check your desktop and download Java and the APPX client if needed, then launch AXAEM. After that, when you click it will just open up. You can choose to make that window password-protected or not.

Versatile handles space management quite well. We know how many boxes can fit on a given shelf and how many spaces are empty. It is a graphical system that will display shelves visually. The only limit is that it prefers to have the same-sized boxes on the same shelf and not mix cubic foot boxes with half-Hollingers. All of our box disposition reports are run from Versatile. The Versatile barcode is produced immediately when boxes are accessioned. When records are processed, we may keep that same barcode, or if boxes are consolidated during processing then they may be re-accessioned into Versatile and the old numbers "destroyed". Our permanent records room still utilizes Versatile barcodes even though the space is managed by the Automated Storage Retrieval

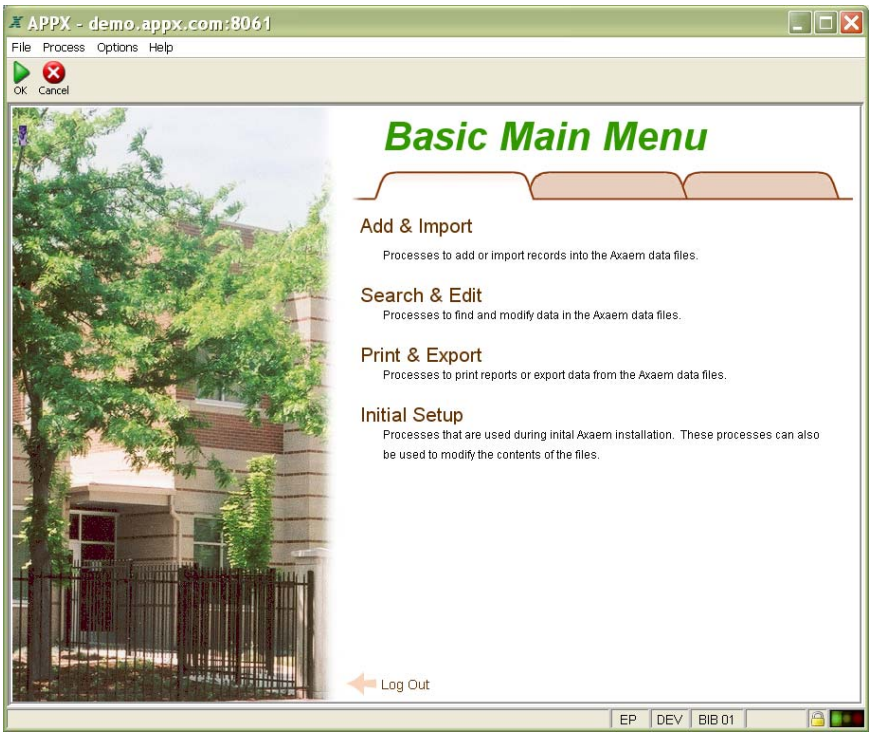
System (ASRS) robots. In this case, the shelf location is removed in Versatile and the storage room field is changed from Records Center to Archives. Boxes are scanned into the ASRS system and then the ASRS database keeps track of where each barcoded box is located so the robots can go out and grab the shelf upon request. In AXAEM, the Research Room staff can request boxes be pulled from the ASRS. You just check off boxes from a list on the screen, and AXAEM generates an email that gets sent to the ASRS handlers, who then make the robot move. We technically could make the robot move directly from AXAEM, but sometimes the robots are down or it would be inconvenient for them to start operating unexpectedly, so we just use email. Once the robot delivers the shelf, a person has to pull the box, put it on a cart, and take it to the Research Room or other staff member anyway. This is how processors request boxes be pulled so they can work on them.

All bibliographic information resides in AXAEM. Versatile is only used to track the location of records and facilitate space management at the records center, including the destruction process. For the records center, it is helpful for them to know what's inside the box in case they need to retrieve something for an agency, and dates are critical for running disposition reports, so generally every box that comes in is given a folder range and date range. That's all that really gets added to Versatile. When we have microfilm, we also add a File record, since multiple microfilm boxes can fit inside one cubic foot box. The File record represents each reel of film, and the only data we add to that record is the microfilm accession number that AXAEM produces. This number is copied into the description field of the Versatile file record. That way, we can locate all the various microfilm copies that are floating around, whether in the ASRS or in the Research Room or even on loan to somebody.

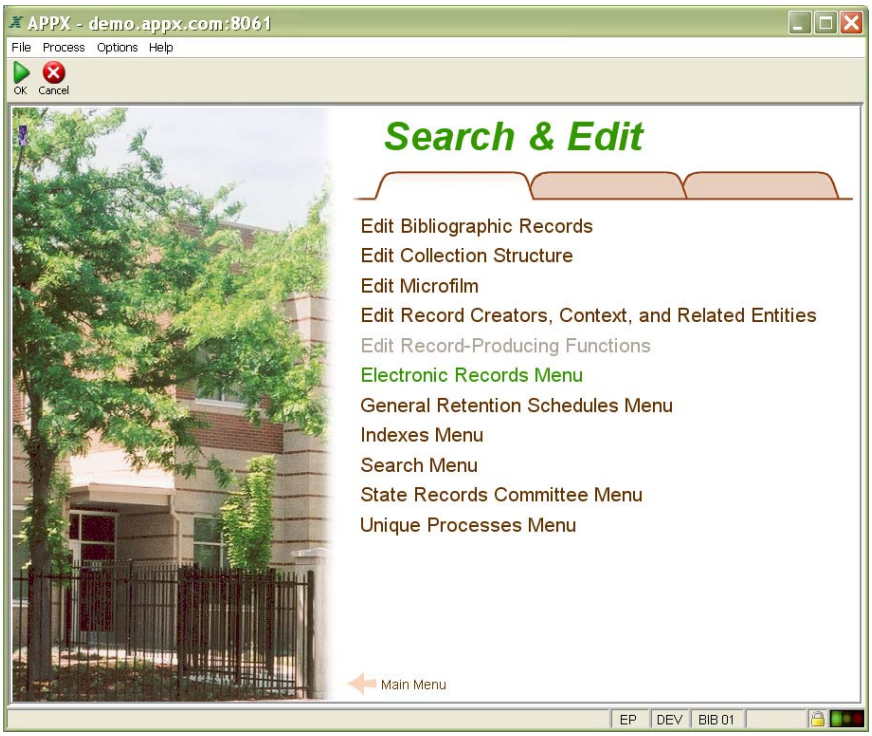
We do have a web form that agencies use to transfer their records to us. It is not directly connected to AXAEM, but could be quite easily. That way there would be no need to scan the transfer documentation.

Appendix 2: AXAEM Testing Instructions

When you login to AXAEM, you will be presented with this menu:

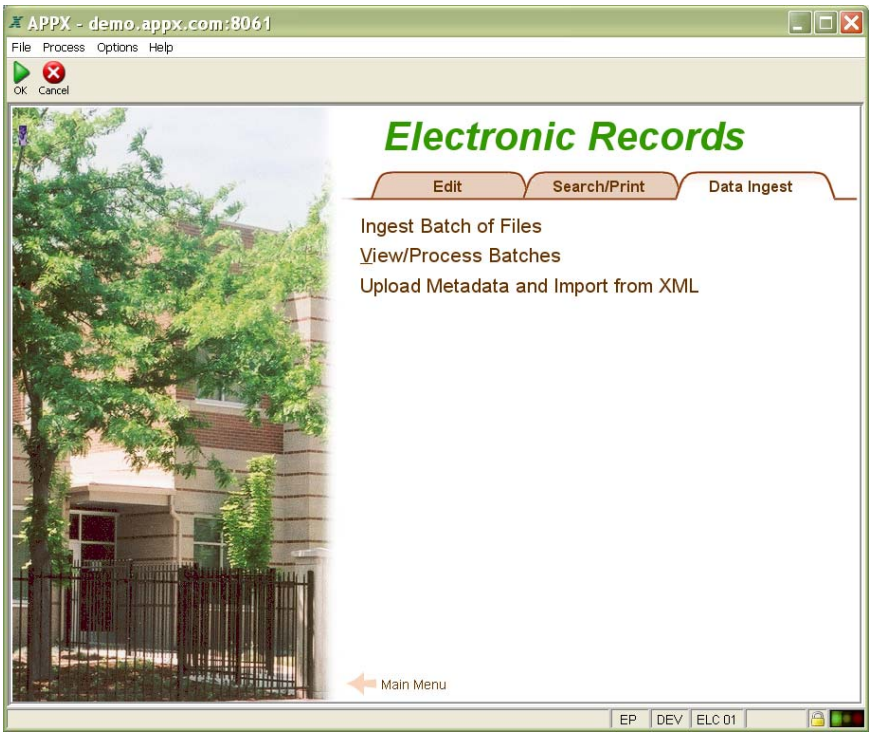


While you can use the Add & Import menu to ingest electronic records, it will probably be easier to work with the Electronic Records menu directly. To get there, choose Search & Edit:

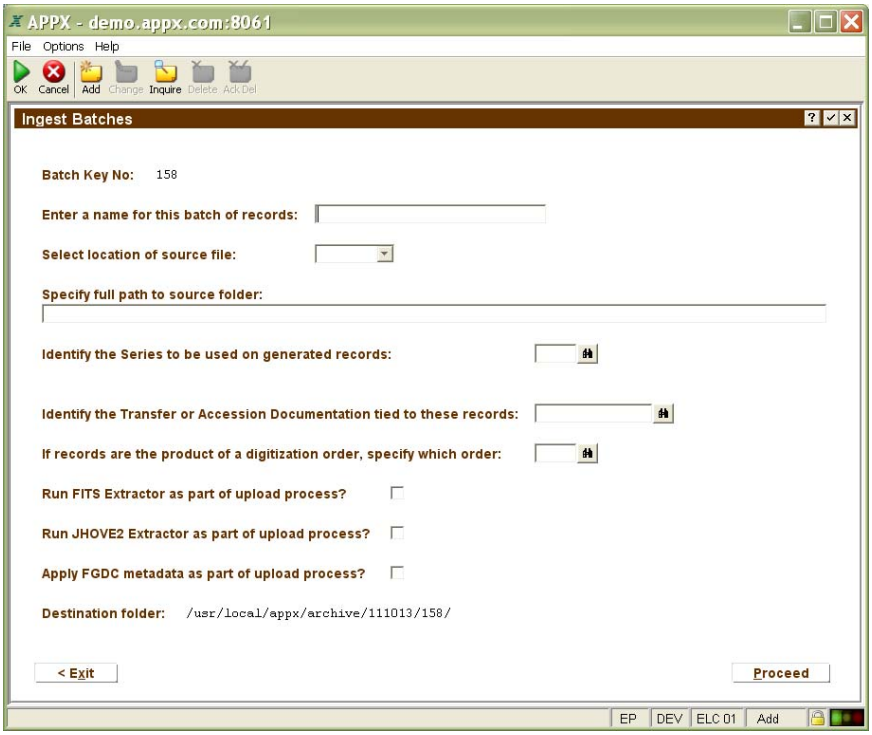


Click on the Electronic Records Menu:

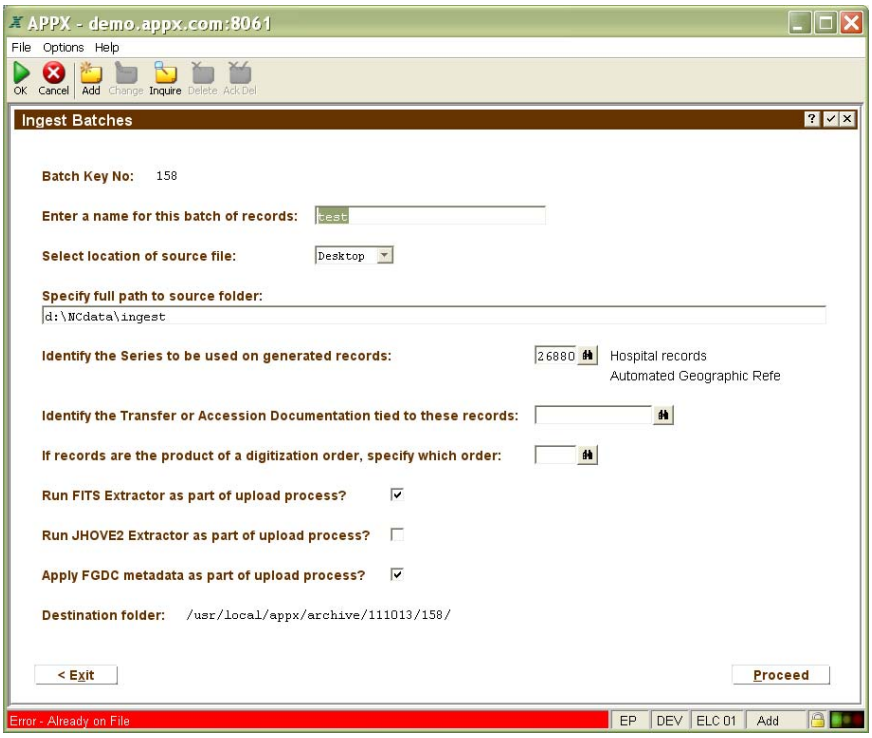
Then go to the Data Ingest tab:



Click Ingest Batch of Files:



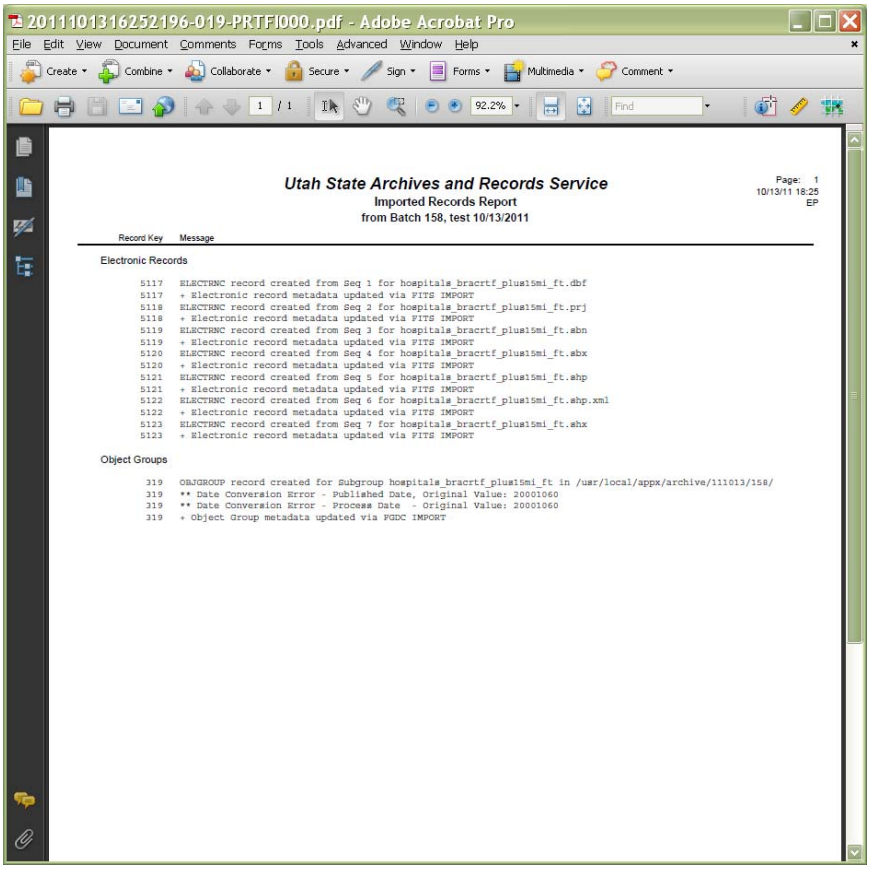
Enter any unique name to identify this batch. In our office we use a combination of the collection/series ID plus today's date. If the name you choose has already been used by another batch, an error will show up on the status bar at the bottom of the screen:



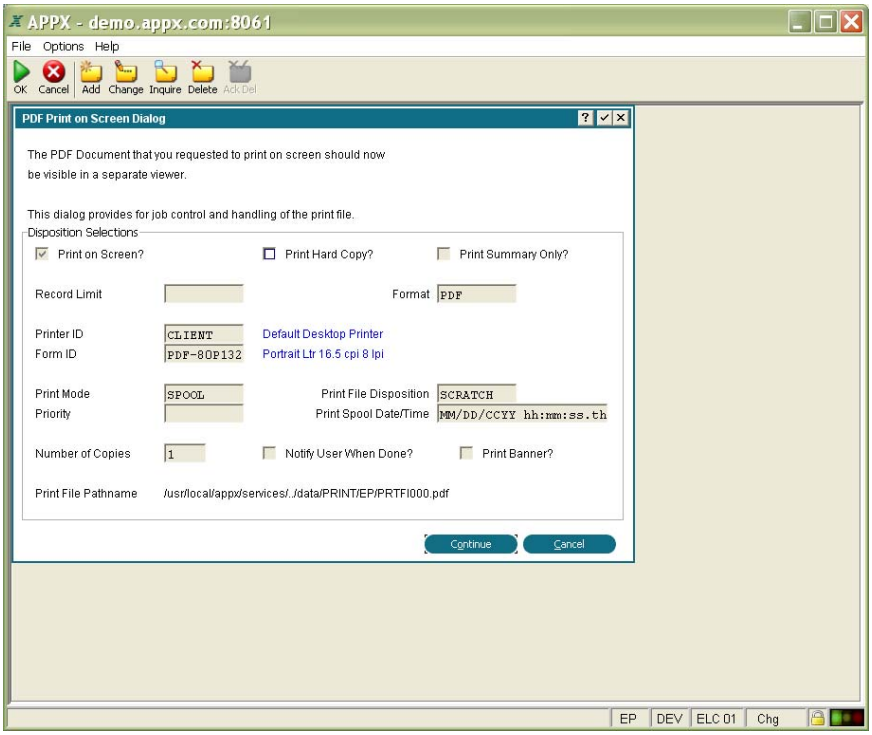
Many of the fields listed are required: name of batch, location of source file, folder being ingested, and series number. You may ignore the links to transfer documentation and digitization orders. You may choose which, if any, metadata extractors to use. I can't guarantee that anything useful will come from JHOVE2, since it's currently being run against individual files instead of whole folders. It's an easy change for us to make it run against the folder, it just hasn't been done yet, and probably won't be for a few days.

For the simplest testing, you may use any random series number that you like. You can click the Scan button to the right of the field to find a likely candidate. If you want to create a series that's more relevant to your records, then please follow the instructions contained in a separate document: How to Add a Collection and a Series. (Hint: go to the Add & Import menu, then Add Collections and Bibliographic Records, then choose Add Bib Records Wizard. In lieu of adding your own creator ID for this collection/series, use any number you like. The number 11 is for the Utah State Archives. After the wizard gets to the end, select the option to edit the series, then use the series number displayed when ingesting electronic records.)

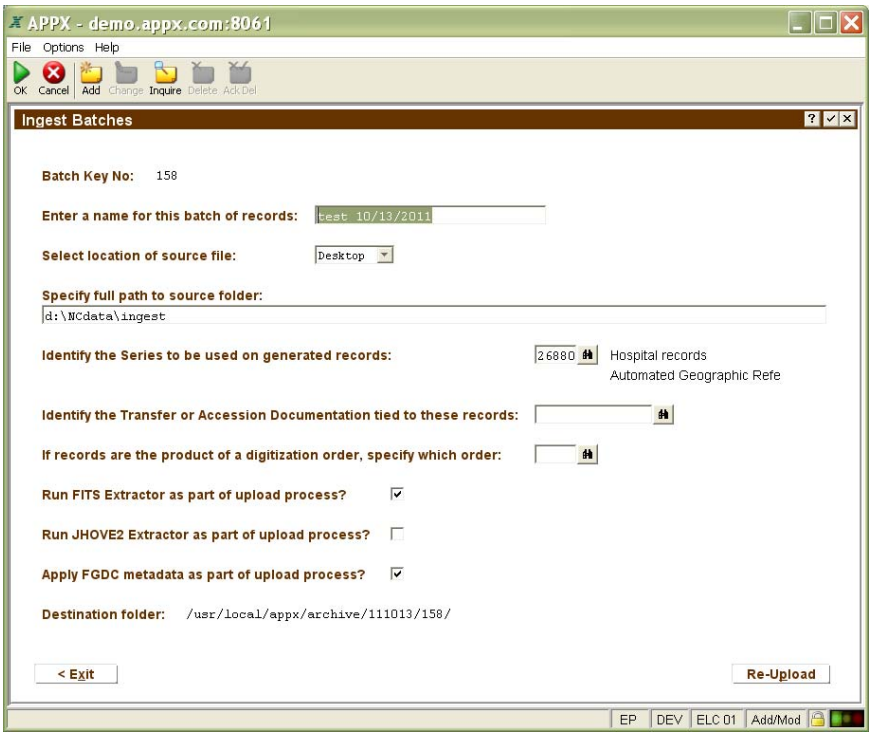
For now, only ingest files which are not zipped. The .zip file will ingest as a single item, but the system won't extract the contents yet. When you are ready to run the ingest, click Proceed. The first click will display the series title information. The second click will run the ingest. After it runs, a PDF report will be generated that lets you know how things went. Keep the report handy for easy reference to the keys to the database files.



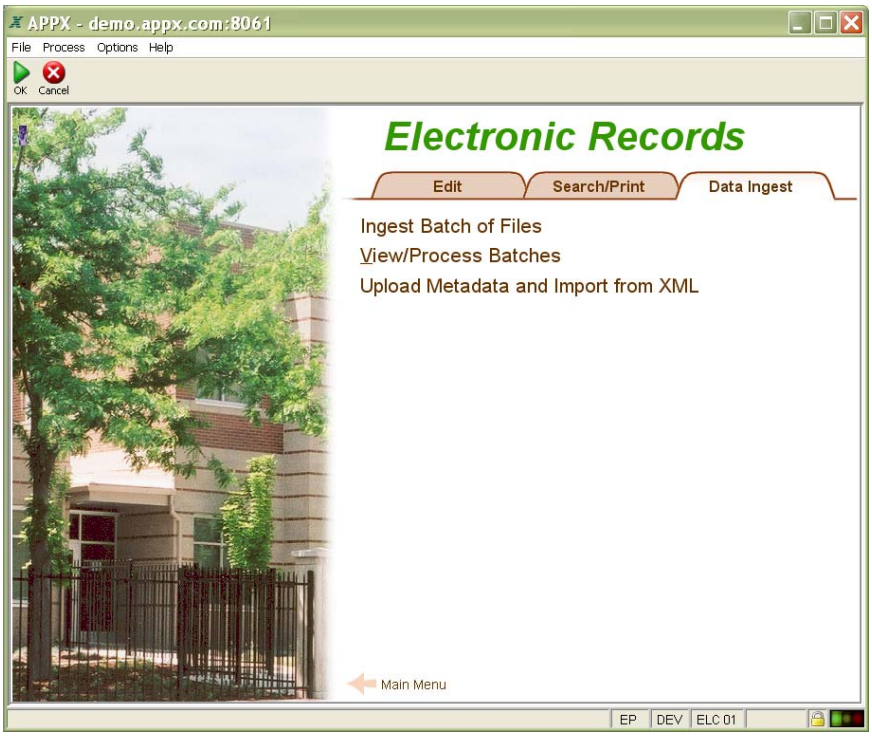
When the following screen comes up, click Continue.



Click Exit from this screen to go back to the menu.

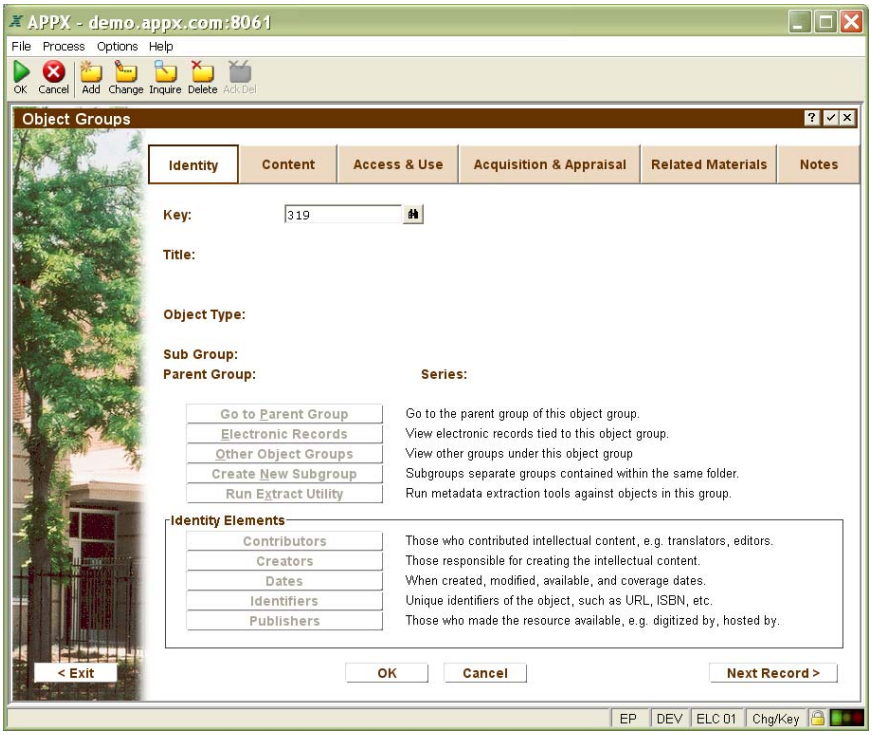


From this menu, click the Edit tab:

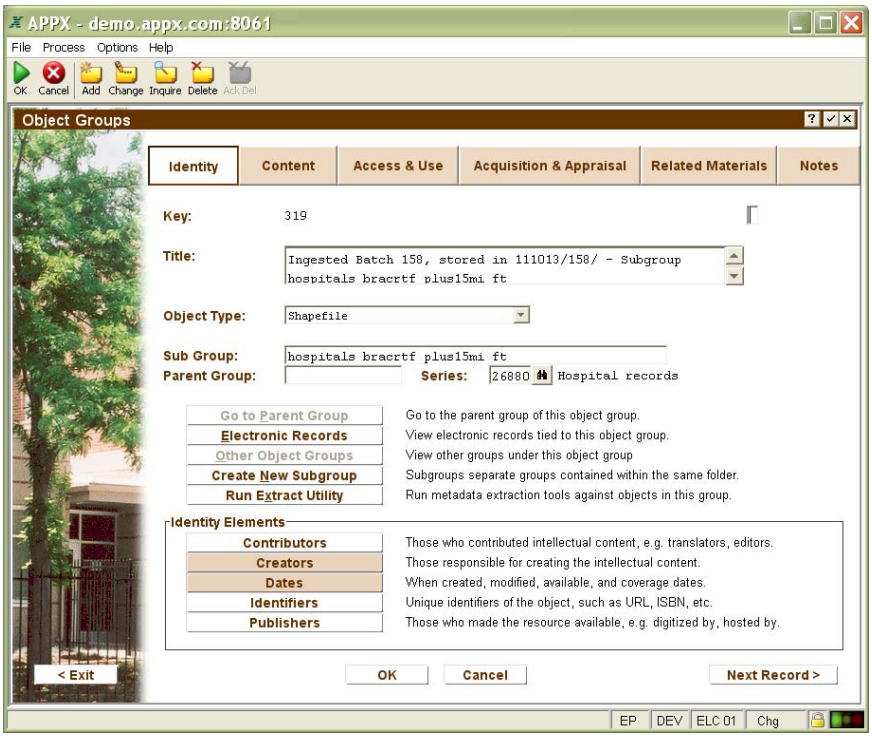


Now you should be back here:

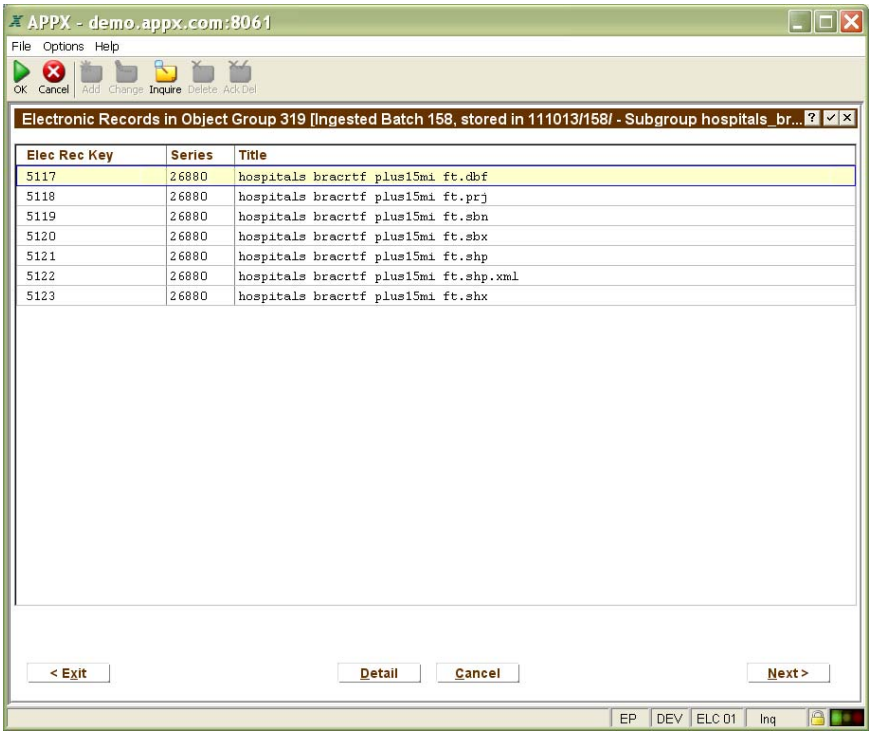
Click the Electronic Object Groups option, then type the key number found in the PDF report:



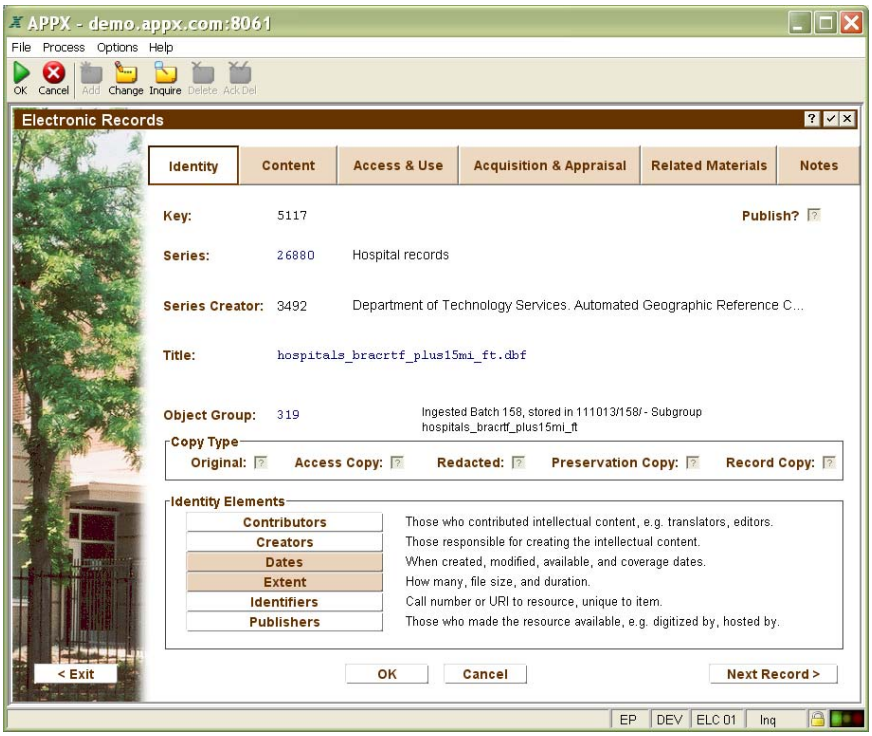
Press Enter or click OK to bring up the record:



You may view the metadata associated with this object group by clicking buttons. Highlighted buttons means there is data there. Other metadata may be viewed from within the other tabs at the top of the screen. To drill down to the electronic records associated with this object, click the Electronic Records button.



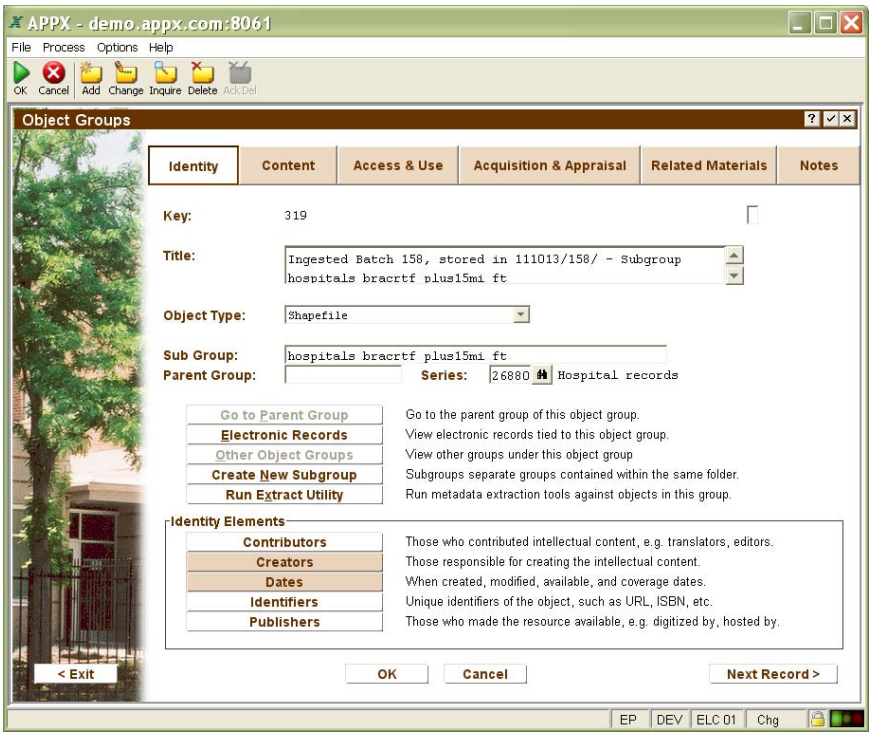
Choose a file and click the Detail button:

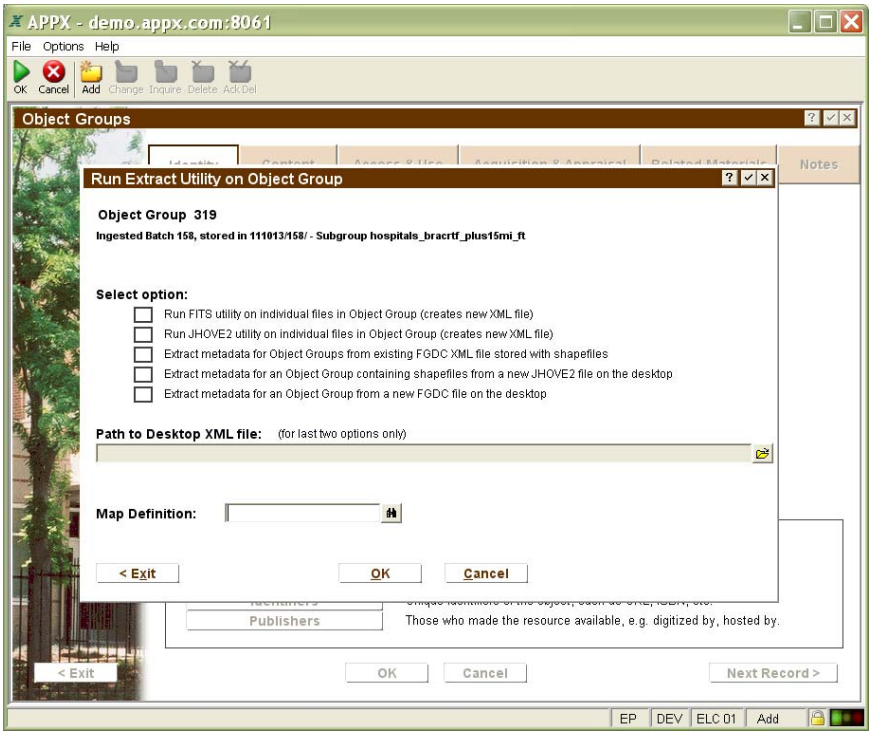


This is the same data you would see if you had chosen to go into the Electronic Records option from the main Electronic Records menu. You may view metadata associated with this file. Descriptive metadata is found in the Content tab. Server location data and the XML file produced by the metadata extractors is found in the Acquisition & Appraisal tab, and the checksum is found in the Notes tab. For shapefiles, descriptive metadata is written to the Object Group fields rather than the Electronic Record fields. Electronic Record data is unique to a physical item, but Object Groups describe the intellectual item.

Click Exit to return to the Object Group.

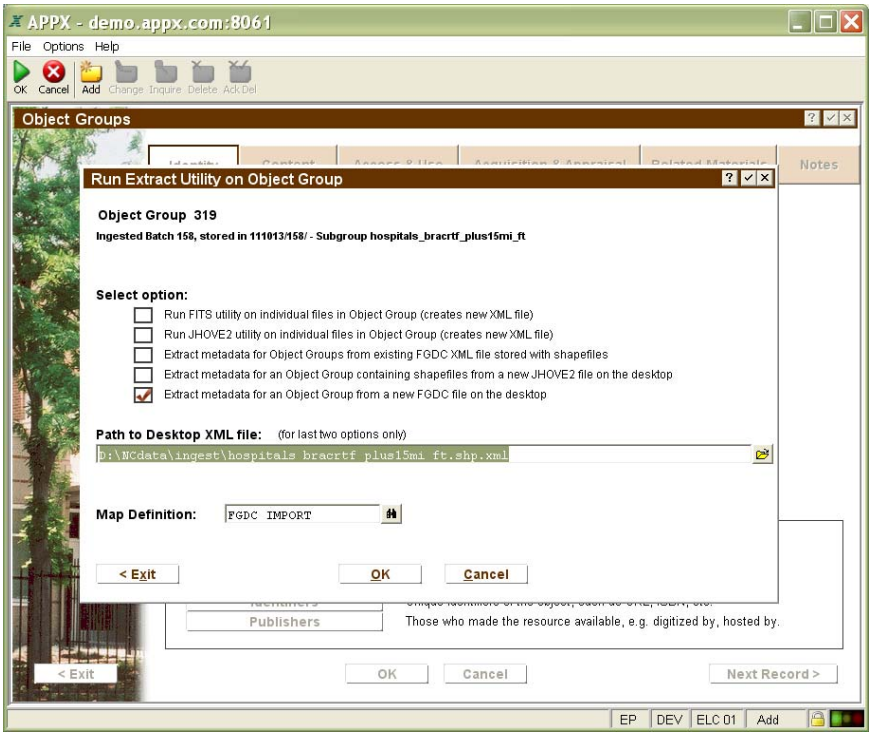
If you did not run a metadata extractor upon ingest, but want to do so after the fact, or you may have upgraded your extractor tool and so want to see the output from the new version, click Run Extract Utility.





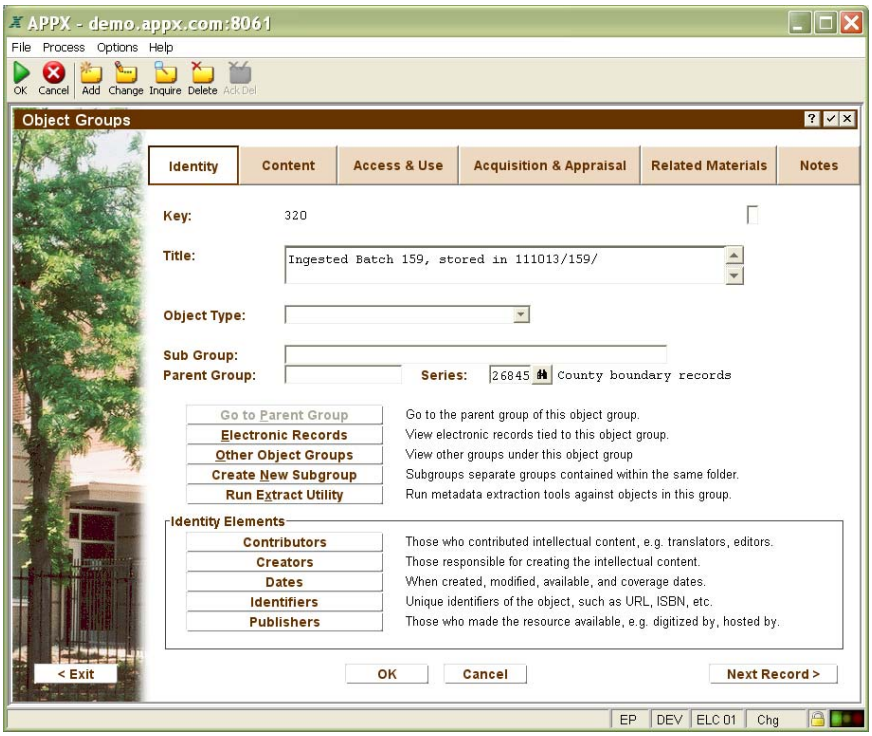
Choose which option you want to run. The last option might be especially useful if you ingested geospatial records that are not shapefiles, but there is FGDC metadata available. Since AXAEM looks for certain patterns of file naming conventions to auto-determine if an XML file is FGDC, and the FGDC file isn't always stored in the folder of the records

it is describing (as is the case with file geodatabases), then you have to populate the metadata after ingest. Click “Extract metadata for an Object Group from a new FGDC file on the desktop,” and provide a path to the FGDC file. Then click OK to ingest the metadata and apply it to the Object Group:

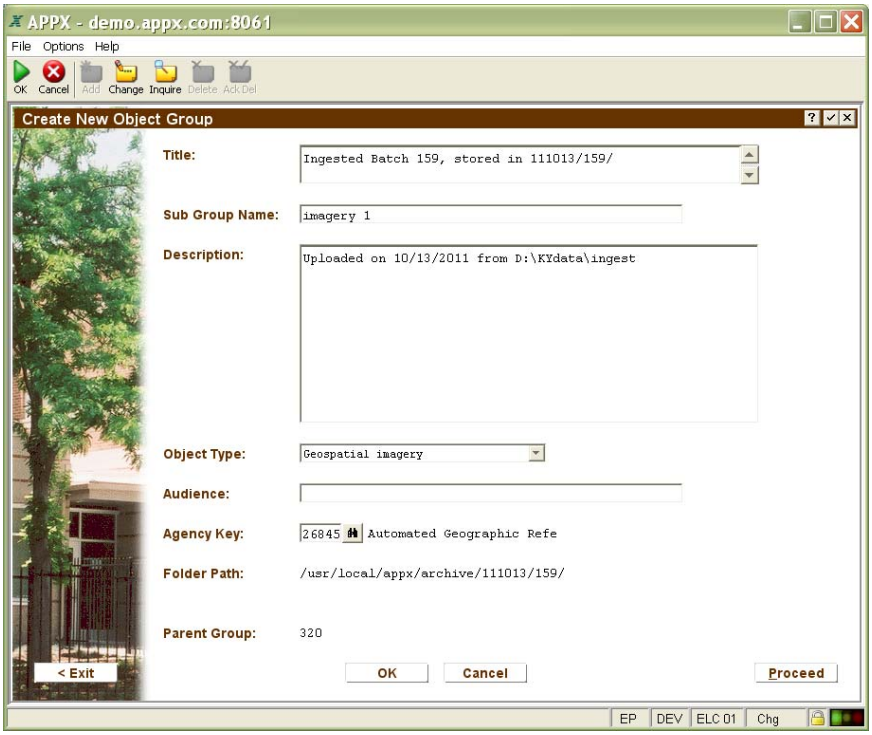


Be sure that the FGDC XML file is actually XML, and not just an HTML file with an .xml extension, since ESRI can produce it that way. Sometimes FGDC may be available for imagery files, too, and a whole folder could hold many imagery/metadata sets. In that case, you would first need to create subgroups within the object group, then apply the right metadata file to the specific subgroup.

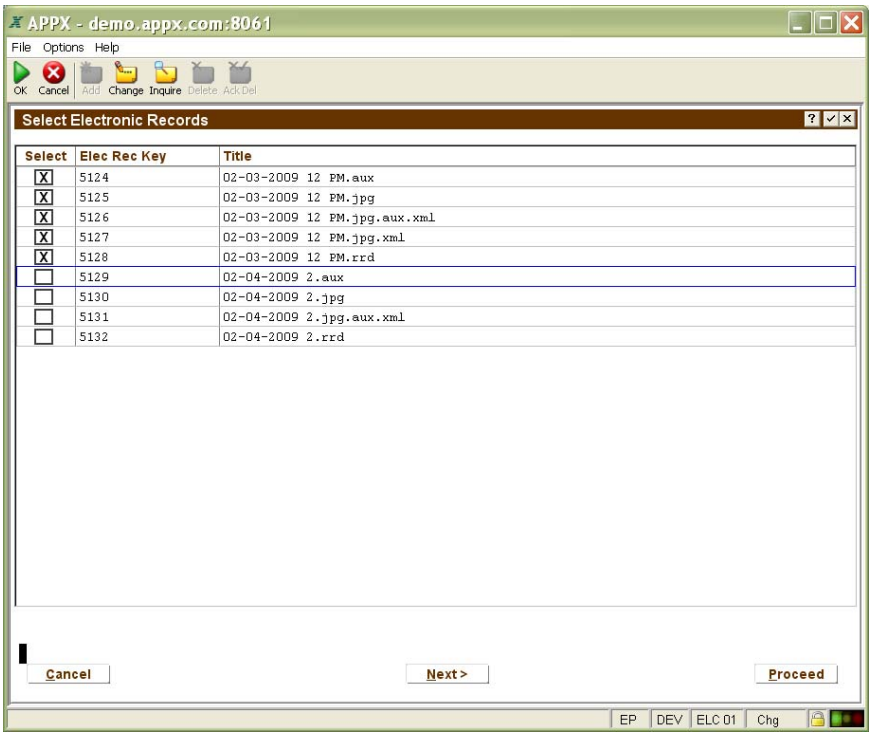
Click on the Create New Subgroup option from the front screen:



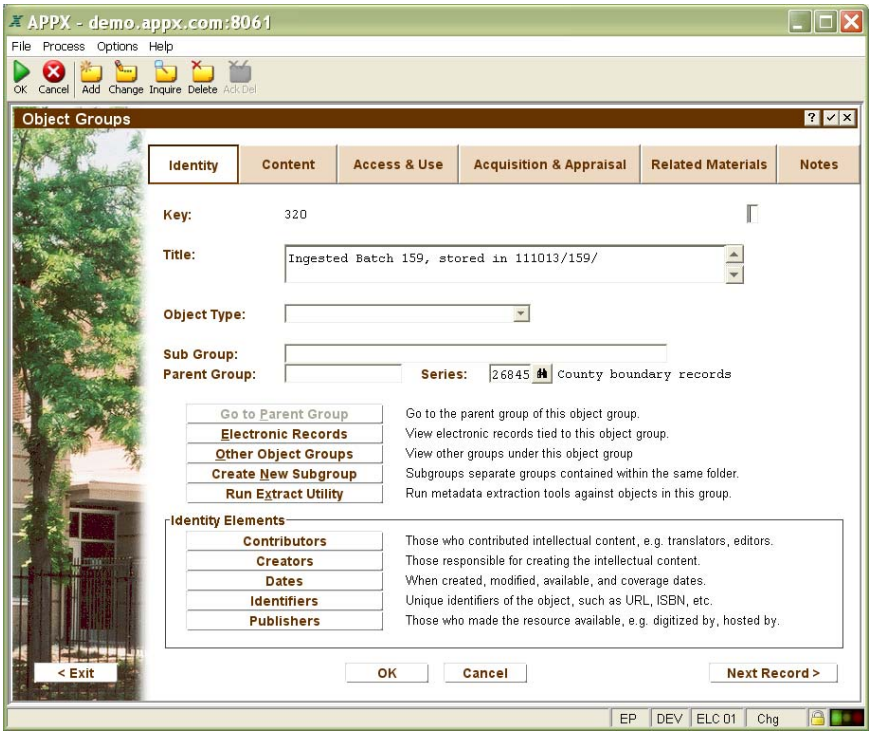
Add a Subgroup name and an object type:

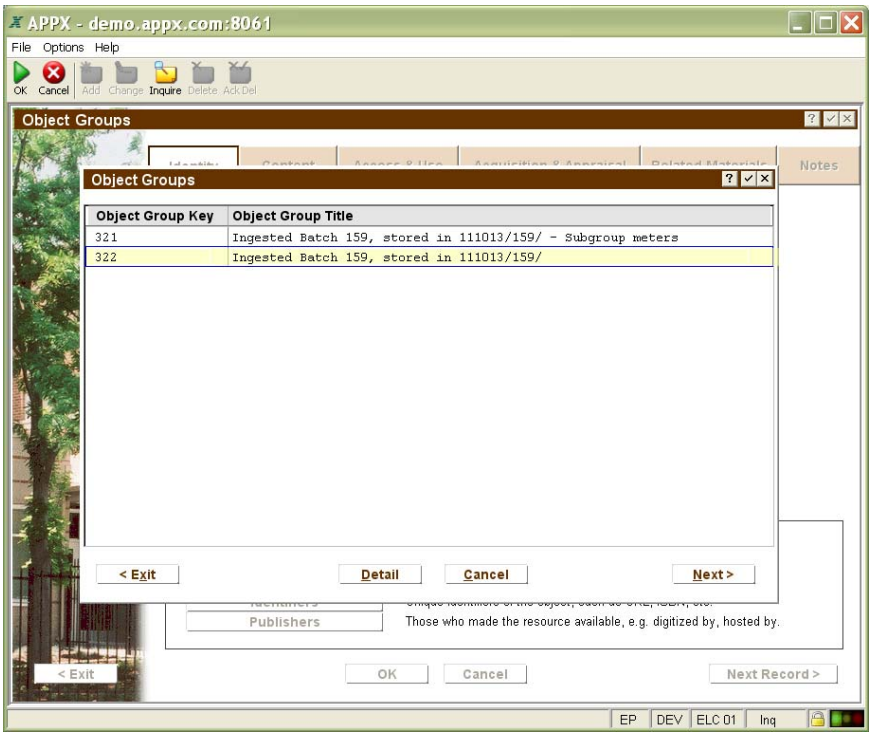


Click OK or Proceed, then check off the records you want in the subgroup:

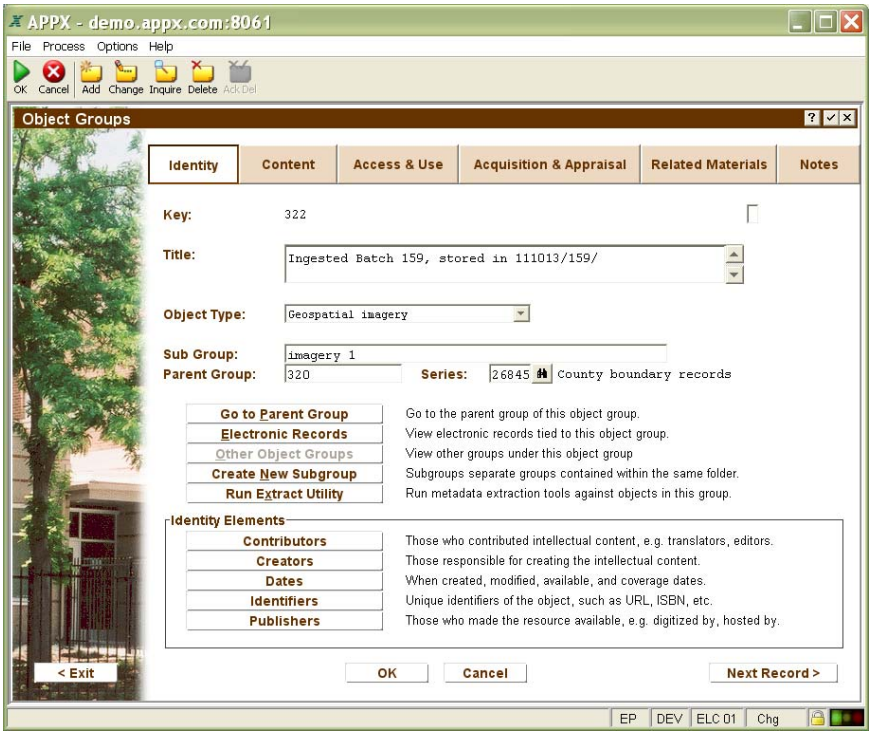


Click Proceed. (If you click Next, it will ask for the Electronic Record ID of a file not listed in the current Object Group.) The system will create the subgroup, then send you back to the first screen so that you can access the group. Click Other Object Groups:

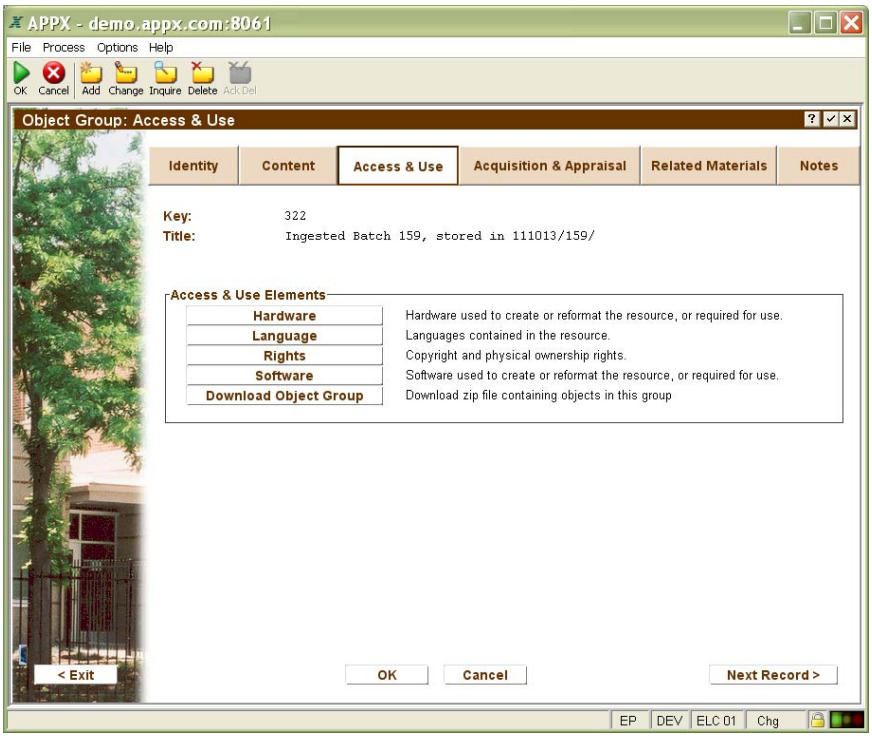




The name of the subgroup should be displayed here (imagery 1). Don't know why it's not. I'll add it to my list of things to fix. Click Detail to see the Subgroup:



To download a zipped copy of these files back to your desktop, click the Access & Use tab:

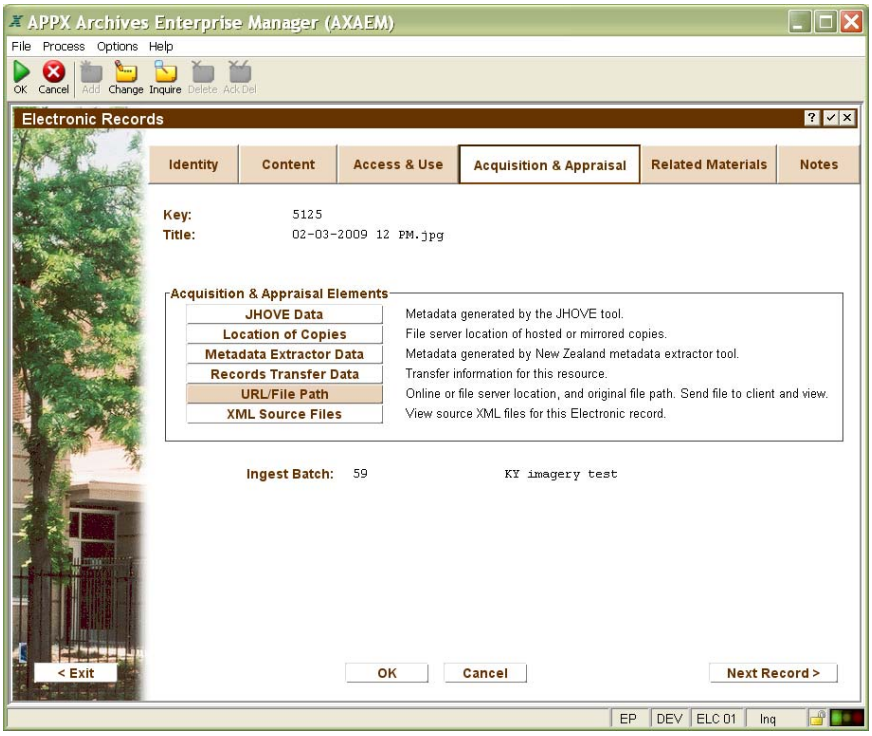


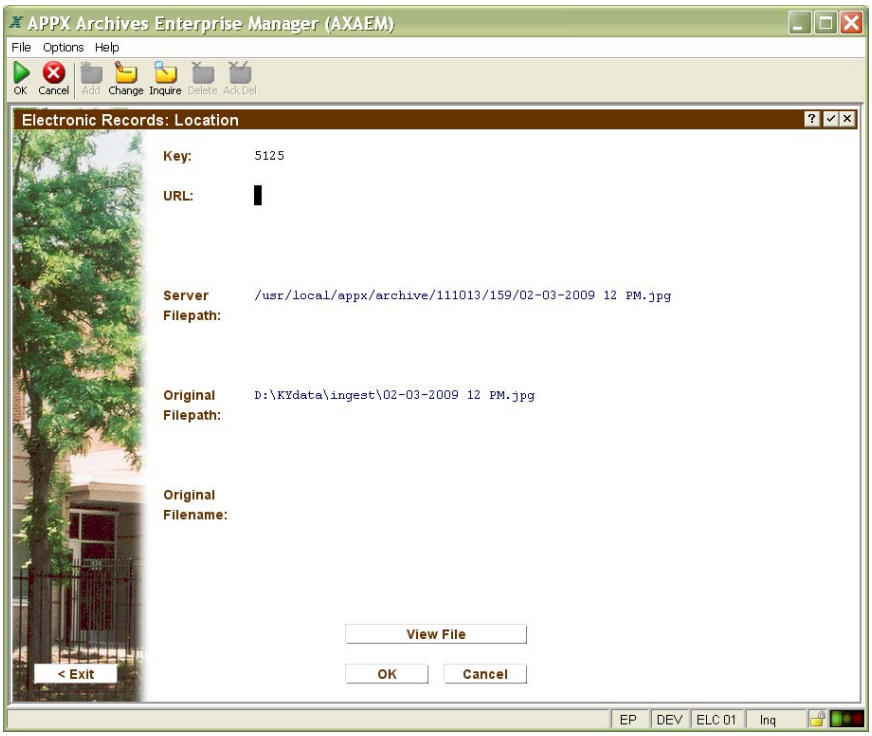
Then click Download Object Group

The .zip file should then open on your desktop in Windows Explorer or other software where it can then be extracted and opened in ArcGIS.

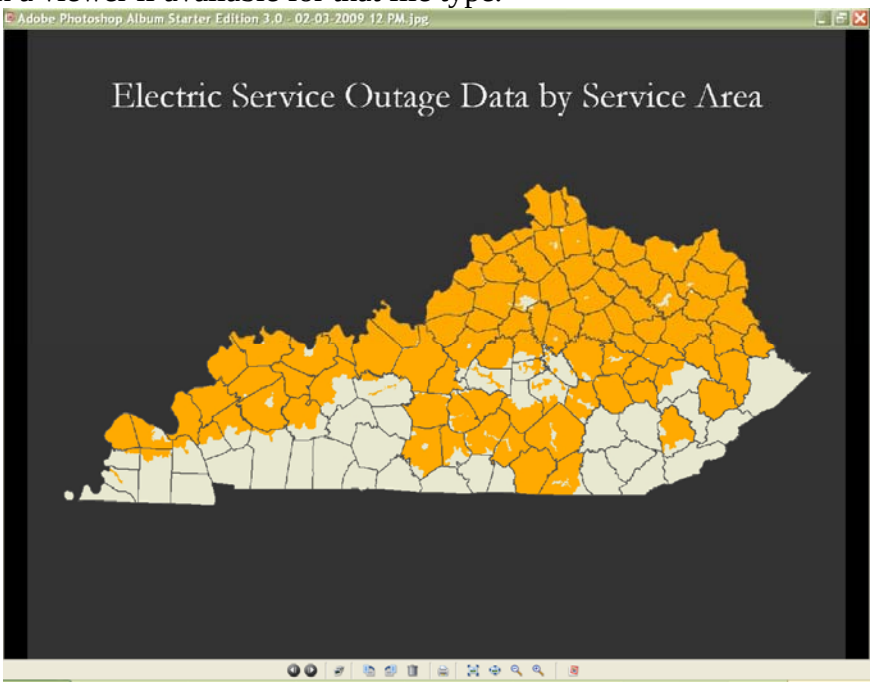
You can also download individual files (unzipped) by going into the Electronic Records option from the menu, pulling up the record in question, clicking the Acquisition & Appraisal tab, then URL/File Path, then View File. This is especially useful when the single file is an image such as a .jpg and you just want to look at it:


APPX - demo.appx.com:8061
File Process Electronic Records Menu Options Help
OK Cancel
Electronic Records
Edit
Search/Print
Data Ingest
Electronic Records
Electronic Object Groups
Format Identification and Migration Rules
Edit Metadata: Export to .csv
Edit Metadata: Import from .csv
Edit Metadata: Templates for Exporting
Hosted Locations of Copies
Individual Media
Media Types
Records Transfers
Main Menu
EP DEV ELC 01


APPX Archives Enterprise Manager (AXAEM)
File Process Options Help
OK Cancel Add Change Inquire Delete AckDel
Electronic Records
Identity
Content
Access & Use
Acquisition & Appraisal
Related Materials
Notes
Key: 5125
Title: 02-03-2009 12 PM.jpg
Acquisition & Appraisal Elements
JHOVE Data
Metadata generated by the JHOVE tool.
Location of Copies
File server location of hosted or mirrored copies.
Metadata Extractor Data
Metadata generated by New Zealand metadata extractor tool.
Records Transfer Data
Transfer information for this resource.
URL/File Path
Online or file server location, and original file path. Send file to client and view.
XML Source Files
View source XML files for this Electronic record.
Ingest Batch: 59 KY imagery test
< Exit
OK
Cancel
Next Record >
EP DEV ELC 01 Inq

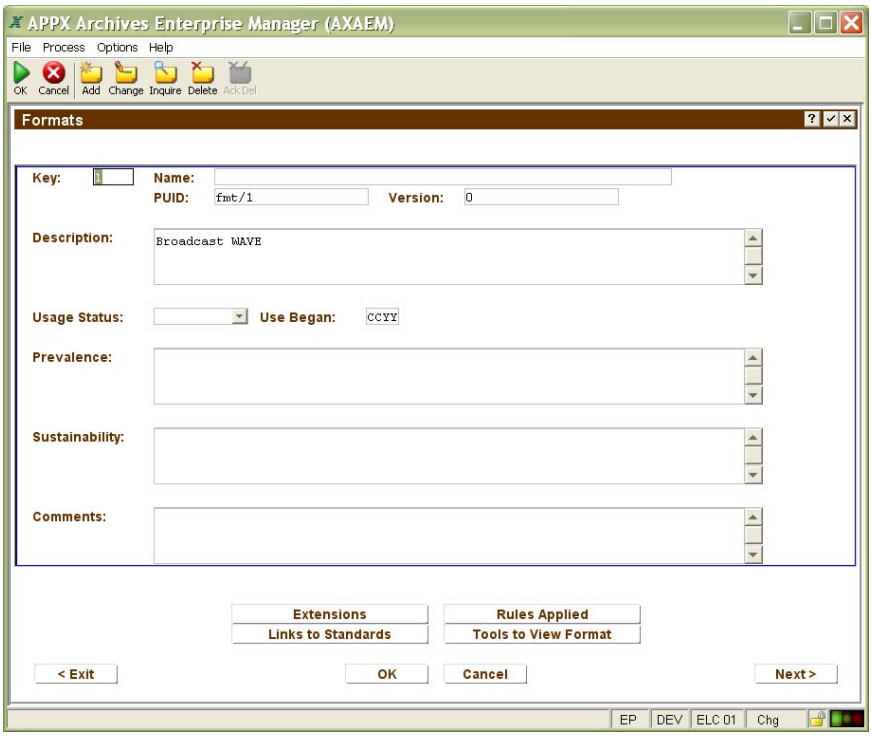


From the above screen, click View File to bring the file down to your desktop, which will auto-launch a viewer if available for that file type.

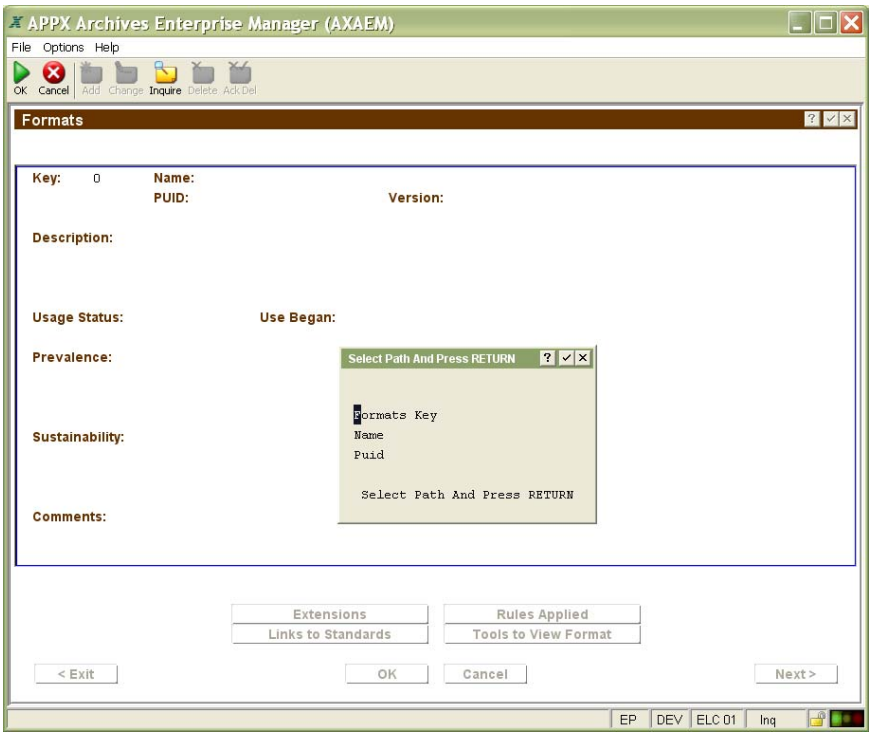


If you like, you may explore other options on the Electronic Records menu, such as Format Identification:

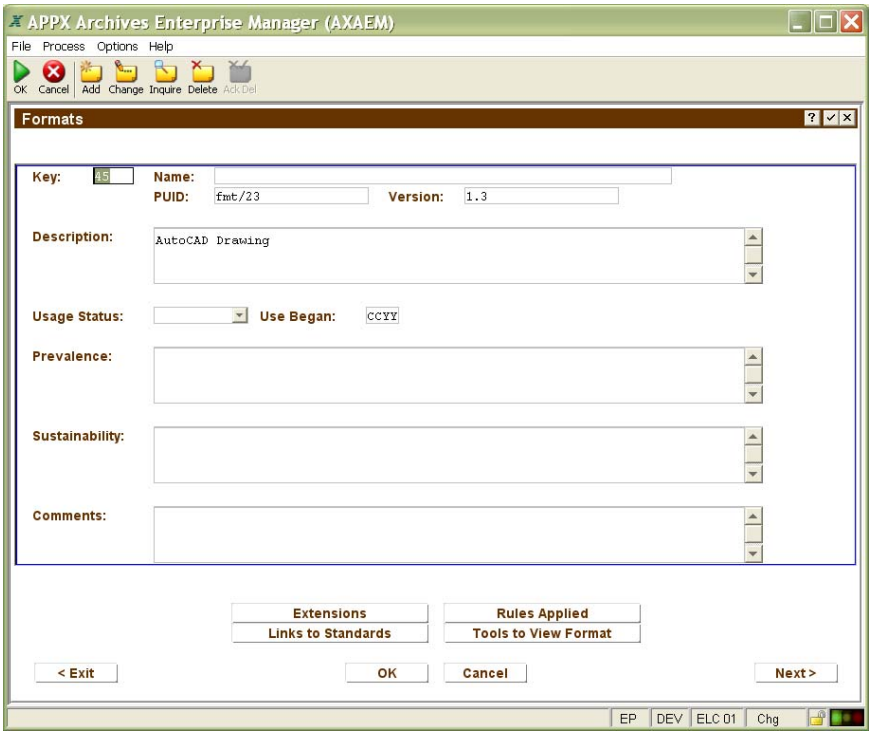
We've imported what's in the PRONOM database, but you can add other formats not yet in PRONOM. Fields were added to store the type of information contained in GeoMAPP's spreadsheet of formats, although the data from that source has not yet been uploaded. This function allows users to identify migration rules per format. Code is not in place yet to actually do any migration. I expect that future development will include a place for users to specify many actions at certain event points, such as what to do upon ingest (how many copies to create, where to put them, plus what to do if a particular format is being ingested).



When you first go into this screen, it displays the first record on the list. To go to a different record, go into Key Entry mode by pressing the Change button on the toolbar (or F12), and then either identifying the key number desired, or selecting a different key path. To select a different key path, from a blank Key Entry screen press F3.



Three choices will come up. To search by PUID, arrow down to that option and press Enter. The Key Entry screen will then allow you to type a PUID that you want to search for. Type a value and press Enter to bring up the record:

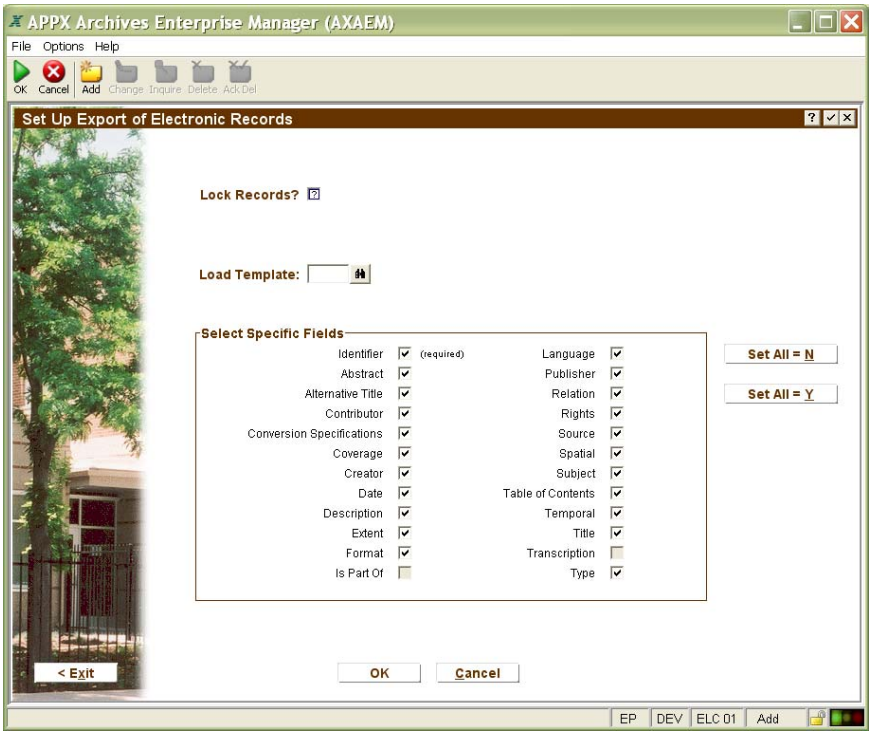


You may add extensions, standards, rules, or tools.

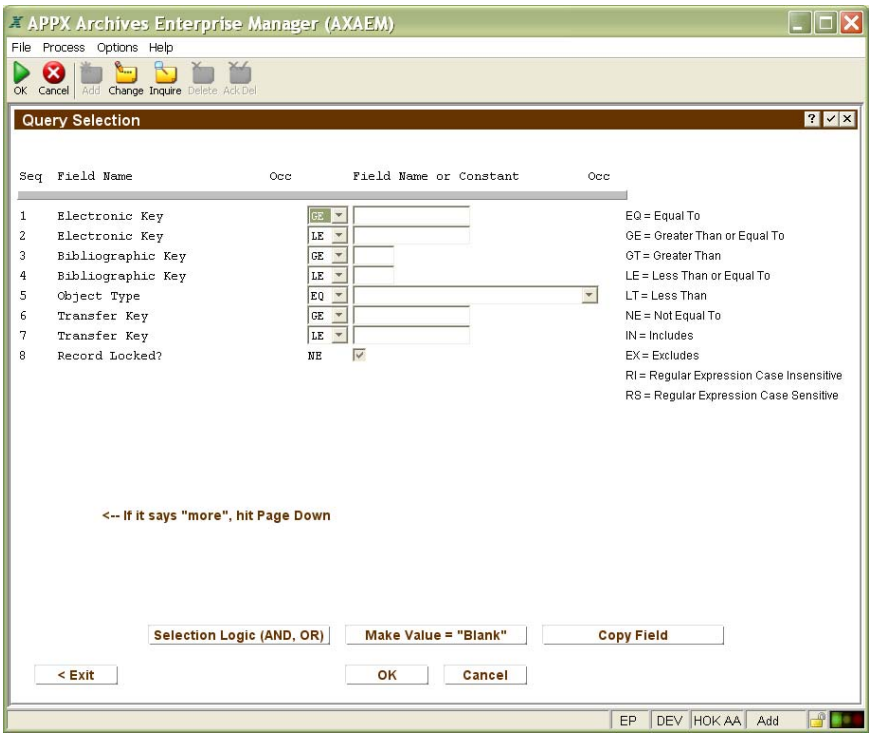
Click Exit to return to the menu.

If you've added metadata to the Electronic Record, you may export it to a .csv file. The .csv may be used to repurpose the metadata in some other system, or be used to help with data entry in this system. If you edit the .csv and then re-import it, the associated records in AXAEM will be updated. Currently this option is only available for metadata associated with an Electronic Record, not an Object Group.

Choose which Dublin Core fields you would like exported:



After you've made your choice, the query will ask which records you want to export. If you don't limit the report by any criteria, then everything in the database will be included in the spreadsheet:



Click OK and the spreadsheet will display on your desktop. Excel displays the data faster if the software is already open and running before you export the metadata.

Click Exit to return to the menu.